

RAMON DE FREITAS SANTOS

**PADRÕES DE POLIADENILAÇÃO EM
MOLÉCULAS DE RNAs DE *Leishmania* OBTIDAS
DA ANÁLISE DE TRANSCRIPTOMA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Agrícola, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

S237p
2012

Santos, Ramon de Freitas, 1983-
Padrões de poliadenilação em moléculas de RNAs de
Leishmania obtidas da análise de transcriptoma / Ramon
de Freitas Santos. – Viçosa, MG, 2012.
xiii, 173f. : il. (algumas col.) ; 29cm.

Inclui anexos.

Orientador: Juliana Lopes Rangel Fietto.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 104-113.

1. Ácido ribonucléico - Síntese. 2. Sequenciamento de
nucleotídeo. 3. *Leishmania*. I. Universidade Federal de
Viçosa. II. Título.

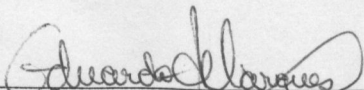
CDD 22. ed. 572.88

RAMON DE FREITAS SANTOS

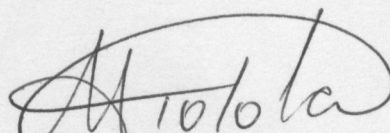
**PADRÕES DE POLIADENILAÇÃO EM MOLÉCULAS DE RNAs DE
Leishmania OBTIDAS DA ANÁLISE DE TRANSCRIPTOMA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Bioquímica Agrícola, para obtenção do título de *Doctor Scientiae*.

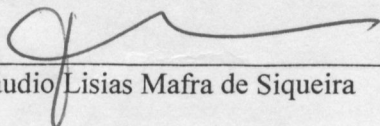
APROVADO: 27 de Junho de 2012



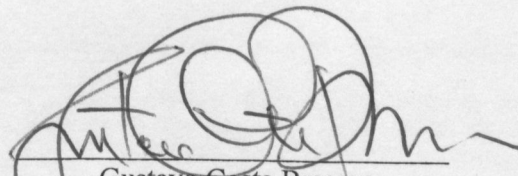
Eduardo de Almeida Marques da Silva



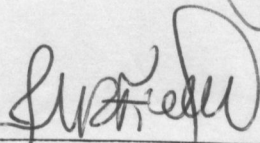
Antônio Helvécio Tótola



Cláudio Lisias Mafra de Siqueira



Gustavo Costa Bressan



Juliana Lopes Rangel Fietto
(Orientadora)

“The imagination is more important than knowledge.”

Albert Einstein

DEDICATÓRIA

Dedico esta conquista aos meus familiares que sempre me apoiaram e incentivaram, em especial ao meu pai Januário Joaquim S. Neto, minha mãe Ivanilda L. Freitas Santos e minha avó Nair Lopes pela preocupação e cuidado. Também não teria chegado a lugar nenhum sem a presença dos grandes amigos em minha vida, vocês foram (e são) muito importantes nesta caminhada, especialmente: Matheus S. Bastos, Ronny F. Souza, André M. Machado, Ronan F. Santos, Leandro S. Santos e Girley Ferreira. Por fim, à Ana Carolina, a pessoa especial que esteve em tempo integral nesses últimos meses me dando forças pra lutar, vencer e conseguir. Ela me fez acreditar que se pode ir muito além, mesmo achando que não se pode mais. Vocês são Especiais e moram no meu coração!

AGRADECIMENTOS

Meus sinceros agradecimentos...

À minha Querida Família, meu Porto Seguro e Ponto de Equilíbrio. Em especial aos meus pais, Januário e Ivanilda, pelo apoio e confiança, meus irmãos, Bruna e Ronan, e meu cunhado “Bola Murcha”, Luciano, pela amizade e incentivo. À minha “vovozinha” Nair, agradeço todos os dias pelo amor, carinho e pelas orações. Amo muito todos vocês!

À Ana Carolina R. Maia, um anjo que “re”apareceu em minha vida. Não imagino a conclusão deste manuscrito sem você ao meu lado, me apoiando. Obrigado por tentar me entender quando eu falava de “poliadenilações” e afins, por ler quase na íntegra a minha tese, por corrigir, comigo, todas as alterações sugeridas. Amo você “Buneca.”

À orientadora, Juliana L. R. Fietto, por aceitar mais este desafio e, acima de tudo, por confiar mais uma vez no meu trabalho. Obrigado pelas palavras de incentivo, por me ouvir nos momentos de desespero (foram vários...), pelas conversas agradáveis do dia-a-dia, pelo exemplo de profissional dedicada e competente. Nesses quase nove anos de convivência aprendi com você, além de fazer ciência propriamente, que ela é realizada com respeito e coletividade. É muito difícil agradecer a você minha querida amiga, as palavras me faltam e os agradecimentos se misturam aos elogios. Minha sincera gratidão pelos ensinamentos, pela amizade e pela credibilidade.

Ao professor Dr. Luís Carlos C. Afonso, pelas discussões intelectuais, auxílio financeiro, pela infraestrutura disponível em seu laboratório e pelo exemplo de profissional dedicado à pesquisa.

Ao Dr. Sérgio Verjovski-Almeida, IQ-USP, pela infraestrutura e suporte financeiro disponibilizados para a realização das etapas de construção e sequenciamento das bibliotecas transcriptômicas. Ainda, pelas valiosas contribuições intelectuais que enriqueceram o trabalho final.

À Dra. Angela Kaysel Cruz, pelas discussões intelectuais que muito auxiliaram nas análises dos resultados e na escrita do paper, produto final deste trabalho. Muito obrigado pelos auxílios técnicos e financeiros que permitiram as análises de sequenciamento realizadas neste trabalho. Agradeço também a oportunidade de trabalho após a conclusão do doutorado.

Ao professor Dr. Luciano G. Fietto, pela amizade, pelas contribuições científicas e suporte financeiro, principalmente nesta etapa final do projeto.

À professora Dra. Márcia R. Almeida, por ter aberto as portas do Laboratório de Infectologia Molecular Animal e pelos auxílios que financiaram minhas viagens técnicas às diferentes instituições colaboradoras para que eu pudesse desenvolver minha pesquisa.

Ao professor Dr. Abelardo Silva Jr. por ter aberto as portas do seu laboratório no Departamento de Medicina Veterinária.

Ao Dr. João Paulo Kitajima pelas orientações iniciais quanto às análises de bioinformática.

Ao Dr. Sérgio Schenkman, da UNIFESP, pela recepção cordial no seu laboratório, e ao Dr. Antônio pela ajuda técnica com os ensaios de microscopia por fluorescência, medida da taxa de tradução por incorporação radioativa de aminoácidos e Western Blotting.

Ao Dr. Samuel Goldenberg, diretor da Instituição Carlos Chagas (Fiocruz/PR), pela agradável recepção em seu laboratório e à Dra. Lysangela Ronalte Alves pela orientação nos ensaios de fracionamento celular em *Leishmania*.

Ao Dr. Ieso de Miranda Castro pelas palavras cordiais durante os encontros pelos corredores do NUPEB/UFOP e pelo empréstimo dos tubos da Ultra-Centrífuga Beckman.

Ao Ronny F. Souza, meu Ilustre Amigo do Peito, meu Irmão Camarada! Com você não tem tempo ruim, sua alegria contagia a todos que estão à tua volta. Você foi importante também nos momentos em que precisei compartilhar as incertezas, as angústias, os planos futuros e as novidades. Valeu meu camarada!

Ao Grande Amigo Matheus S. Bastos, amizade construída desde a época de graduação em Ouro Preto. Sua ajuda nas etapas finais foi imprescindível para a concretização deste projeto. Obrigado por escutar minhas conversas fiadas, por perdoar meus inúmeros atrasos, pelas parcerias nos botecos e por estar sempre disposto a ajudar. Saiba que admiro muito o seu conhecimento, um dos melhores e mais completos profissionais com quem tive a oportunidade de trabalhar.

Ao meu Grande Amigo Breno de Assis, pela amizade e pelas inúmeras risadas proporcionadas pelas conversas malucas, pelas caminhadas na Linda UFV, não como atividade física propriamente, mas muito mais por lazer, pelo simples prazer de trocar ideias superinteressantes e muito viajadas!

Ao Girley Ferreira, amizade construída desde a época de República na saudosa Terrinha de Ouro Preto. Obrigado pela força em São Paulo, as várias hospedagens no seu “apê” foram indispensáveis quando a grana era curta.

Aos Amigos de Ouro Preto, em especial, Leandro S. Santos, José Luiz Rocha e André M. Machado, três irmãos que considero. A amizade de vocês não tem preço!

Ao Francisney Félix, mais uma amizade construída em Viçosa, parceiro de laboratório, adversário dos “baum” das várias rodadas de sinuca e cerveja às quintas-feiras. Ralamos muito, discutimos muita bioquímica durante o cumprimento dos créditos (tenso...), época boa aquela viu.....Um Grande Amigo é pra sempre!

Ao grande Felipe F. Castro, o melhor e mais dedicado aluno de Iniciação Científica que o LIMA já teve, deixo aqui registrado meus sinceros agradecimentos pelas ajudas no LIMA, por botar o pé na estrada e encarar o trabalho na UFOP, pela

disposição em fazer ciência todas as vezes que precisei de ajuda, pelas trocas de ideias viajadas, pelas parcerias nas baladas, pela camaradagem em Ribeirão Preto e, principalmente, pela amizade.

Ao Rafael Bueno, parceiro de pós-graduação e Grande Amigo que conquistei durante essa passagem em Viçosa.

Aos Amigos Talles Eduardo F. Maciel e Pedro Marcos P. Vidigal, sem vocês este projeto não teria se tornado real. Muito obrigado pelos auxílios computacionais, este trabalho é tão de vocês quanto meu. Ninguém é tao bom que se possa trabalhar sozinho nem tão ruim que não se possa trabalhar em grupo. Um “salve” à coletividade!

Aos doutorandos da USP, Murilo Sena Amaral e Felipe C. F. Beckedorff, pelo auxílio durante a construção e sequenciamento das bibliotecas transcriptômicas.

Aos colegas do Laboratório de Imunoparasitologia do NUPEB/UFOP pelo convívio agradável e pelo clima “alto astral” durante a jornada de trabalho.

Aos colegas do Laboratório de Infectologia Molecular Animal do BIOAGRO/UFV por tornar mais divertido e alegre os momentos de trabalho.

Ao Sr. Valdir pelas inúmeras ajudas no preparo dos materiais e pelos momentos de descontração!

Às agências financiadoras deste projeto: CAPES, CNPq, FAPEMIG, FAPESP, Agência Financiadora de estudo e projetos (FINEP) e Ministério da Ciência e Tecnologia (MCT).

Ao Ensino Público, gratuito e de qualidade.

Ao Departamento de Bioquímica e Biologia Molecular e ao Programa de Pós-Graduação em Bioquímica Agrícola da UFV pela oportunidade.

Ao Eduardo Monteiro, secretário do Programa de Pós-Graduação em Bioquímica Agrícola, pela atenção e cordialidade em nos receber, sempre com muita dedicação, boa vontade e eficiência para resolver nossos problemas.

SUMÁRIO

LISTA DE FIGURAS E TABELAS	ix
RESUMO	xi
ABSTRACT.....	xiii
INTRODUÇÃO	1
REVISÃO BIBLIOGRÁFICA	6
<i>Leishmania</i> e Leishmanioses	7
Arquitetura e organização genômica em <i>Leishmania</i> spp.....	18
Expressão gênica diferencial no gênero <i>Leishmania</i>	20
Poliadenilação estável da extremidade 3' dos mRNAs em eucariotos	23
OBJETIVOS	27
Objetivo Geral.....	28
Objetivos Específicos.....	28
MATERIAIS E MÉTODOS	29
Camundongos	30
<i>Leishmania</i> spp., Isolados e Manutenção das Populações do Parasito	30
Curva de Crescimento, Inóculo e Acompanhamento do Desenvolvimento de Lesão em Camundongos C57BL/6.....	31
Fracionamento Celular e Purificação da Fração Polissomal	31
Ensaio de Atividade Enzimática da Glicose-6-Fosfato-Desidrogenase.....	32
Extração de RNA Total.....	33
Extração de RNAs poli(A) ⁺	34
Quantificação de RNA e DNA por Método Fluorescente	35
Eletroforese Microcapilar das Amostras de RNA Total e RNA poli(A) ⁺	36
Reação de Transcriptase Reversa [oligo(dT) RT-PCR] e Reação em Cadeia da Polimerase (PCR)	37
Purificação de DNA (amplicons RT-PCR) do Gel de Agarose	38
Clonagem dos Produtos de PCR Purificados.....	39
Preparo de Células Competentes por Cloreto de Cálcio	39
Transformação das Células Competentes por Choque Térmico	40
Miniprep de DNA Plasmidial em Placa de 96 Poços.....	40
Reação de Sequenciamento, Precipitação e Sequenciamento Automático do DNA	42
Tabela 1. Lista de Oligonucleotídeos Utilizados Neste Trabalho.....	43
Construção e Sequenciamento das Bibliotecas Transcriptômicas de RNA-Seq	44

Análises de Bioinformática.....	44
1) Mapeamento de rRNAs.....	44
2) Mapeamento das subunidades do exossoma de degradação de RNAs	46
RESULTADOS.....	47
Co-Purificação e Detecção de Diferentes Subunidades de rRNAs Entre os mRNAs poli(A) ⁺ de <i>Leishmania</i>	48
Análises das Bibliotecas de RNA-Seq: Conjunto de Dados das Sequências de rRNAs em Relação às Bibliotecas de Modo Geral	52
Poliadenilação de Moléculas de rRNAs em <i>L. (V.) braziliensis</i> e <i>L. (L.) amazonensis</i>	56
Poliadenilações Internas em rRNAs e mRNAs Sugerem um Novo Processamento Pós-Transcricional de RNAs em <i>Leishmania</i>	71
Transcritos de rRNAs Internamente Poliadenilados e não Abundantes são Confirmados por Oligo(dT) RT-PCR	74
Fragmentos Truncados Poliadenilados das Moléculas do rRNA LSU 28S beta e do mRNA do eiF5a2 Estão Presentes Tanto no Núcleo Quanto no Citoplasma de <i>L. (V.) braziliensis</i>	82
Proteínas do Complexo Exossoma Estão Presentes no Genoma de <i>Leishmania</i> e são Ativamente Transcritas	86
DISCUSSÃO	88
CONCLUSÕES E CONSIDERAÇÕES FINAIS	101
REFERÊNCIAS BIBLIOGRÁFICAS	104
ANEXOS	114
Figura suplementar S1. Padrão de eletroforese capilar das amostras de mRNAs poli(A) ⁺ isoladas por afinidade oligo(dT) mostram a presença de subunidades de rRNAs em diferentes formas evolutivas de <i>L. (V.) braziliensis</i> e <i>L. (L.) amazonensis</i>	115
Texto Suplementar S1. <i>Reads</i> de rRNAs presentes nas bibliotecas transcriptômicas apresentam caudas homo- ou heteropoliméricas nas extremidades 3'-terminal.....	116
Lbr_ET_MET	116
Lbr_NSL_MET.....	123
Lbr_ET_PRO	137
Lbr_NSL_PRO	145
La_PH8_MET.....	155
Texto Suplementar S2. <i>Reads</i> de mRNAs presentes nas bibliotecas transcriptômicas apresentam caudas homo- ou heteropoliméricas nas extremidades 3'-terminal.	163
Soluções e Tampões.....	167
Água DEPC.....	167
Solução 0.9% NaCl.....	167

Cicloheximida	167
Heparina (20 mg/mL).....	167
Solução de lise NaOH/SDS.....	167
Tampão de lise hipotônico	168
Nonidet P-40	168
Sacarose 2M.....	168
Tampão TAE 10X.....	169
Meio Grace's base.....	169
Meio Grace's completo.....	170
Meio RPMI	170
Meio DMEM.....	170
Gel de agarose 1.2%	171
Ribonucleoside Vanadyl Complex (inibidor de RNases)	171
L-Glutamina (solução estoque 200 mM)	171
Meio LB-ágar.....	171
PBS	172
Tampão TBE 5X.....	172
Tampão GTE.....	173
Penicilina (solução estoque 6 mg/mL).....	173
Ampicilina (solução estoque 5 mg/mL).....	173

LISTA DE FIGURAS E TABELAS

Figura 1. Avaliação da virulência dos isolados ET e NSL de <i>L. (V.) braziliensis</i> em modelo murino (C57BL/6).....	48
Figura 2. Padrão de eletroforese capilar das amostras de mRNAs poli(A) ⁺ isoladas por afinidade oligo(dT) mostram a presença de subunidades de rRNAs.....	50
Figura 3. Representação esquemática do mapeamento do “cluster” de RNAs ribossomais de <i>L. (V.) braziliensis</i> e <i>L. (L.) amazonensis</i>	55
Figura 4. Exemplos de fragmentos de rRNAs com caudas ricas em poli(A) em <i>Leishmania</i>	70
Figura 5. Poliadenilação “canônica-like” e poliadenilação truncada no 28S rRNA beta e no mRNA do eiF5a2 podem ser detectadas por oligo(dT) RT-PCR.....	75
Figura 6. Sítio de poliadenilação interna na subunidade beta do rRNA 28S ribossomal (<i>read</i> GPST6G002H92OB) e no mRNA do eiF5a2 (<i>read</i> GKLMG1101AS3D2) nas bibliotecas transcriptômicas de <i>L. (V.) braziliensis</i>	77
Figura 7. Representação esquemática do alinhamento de <i>reads</i> de rRNA 28S beta mostra dois conjuntos de <i>reads</i> alinhando no “template” LSU 28S-beta do primeiro cluster de rRNAs.....	78
Figura 8. Moléculas de rRNAs 28S e 18S truncadas em <i>L.(V.) braziliensis</i> com caudas poli(A) ⁺ homo- ou heteropoliméricas.	80
Figura 9. Moléculas de mRNAs (cDNA) truncadas em <i>L. (V.) braziliensis</i> com caudas poli(A) ⁺ homo- ou heteropoliméricas nas mensagens da proteína 60S ribossomal L5 e do fator de tradução eiF5a2.....	81
Figura 10. Quantificação do DNA e dosagem da atividade enzimática da glicose-6-fosfato-desidrogenase (G6PDH) nos compartimentos celulares de <i>Leishmania</i> após fracionamento celular.....	83
Figura 11. Eletroforese em gel de agarose 1,2% corado com brometo de etídio dos fragmentos poliadenilados amplificados por oligo(dT) RT-PCR das frações celulares dos promastigotas totais do isolado NSL de <i>L. (V.) braziliensis</i>	85
Tabela 1. Lista de Oligonucleotídeos Utilizados Neste Trabalho.....	43

Tabela 2. Quantificação pelo Bioanalyzer – RNA 6000 Pico (teste de ligação).	51
Tabela 3. Resumo dos dados de sequenciamento das bibliotecas transcriptômicas para <i>L. (V.) braziliensis</i> e <i>L. (L.) amazonensis</i> . Confirmação do alto conteúdo de rRNAs após a purificação dos RNAs poli(A) ⁺ por afinidade com beads magnéticos oligo(dT).	52
Tabela 4. Moléculas de rRNA SSU 18S, LSU 5.8S e LSU 28S rRNA poliadeniladas mapeadas nos dados de transcriptoma em <i>Leishmania</i> spp.	57
Tabela 5. Moléculas de mRNAs truncadas poliadeniladas mapeadas através das análises do dados do transcriptoma para os genes codificantes de proteínas em <i>L. (V.) braziliensis</i>	73
Tabela 6. Genes ortólogos do complexo do exossoma identificados nos dados de transcriptoma de <i>L. (V.) braziliensis</i> (isolados ET e NSL) e no TriTrypDB.	87

RESUMO

SANTOS, Ramon de Freitas, D.Sc., Universidade Federal de Viçosa, Junho de 2012. **Padrões de poliadenilação em moléculas de RNAs de *Leishmania* obtidas da análise de transcriptoma.** Orientadora: Juliana Lopes Rangel Fietto. Co-orientadores: Márcia Rogéria de Almeida, Luís Carlos Crocco Afonso e Luciano Gomes Fietto.

Os sequenciamentos de próxima geração (NGS) em larga escala e de alto rendimento revelaram novos aspectos nos genomas e transcriptomas que não foram observados ou foram pobremente descritos usando sequenciamento em pequena escala. Usando duas espécies de *Leishmania* como modelo numa análise global do transcriptoma por pirosequenciamento (454/Roche) encontramos poliadenilações não canônicas em moléculas de rRNAs e mRNAs. Os RNAs em *Leishmania* são transcritos como longas moléculas policistrônicas pós-transcricionalmente processadas, mas usualmente somente o mRNA é poliadenilado estavelmente na sua extremidade 3'-madura. Analisando os dados de transcriptoma também observamos caudas poli(A)⁺ ou ricas em poli(A)⁺ na extremidade 3' dos genes de rRNA. Este fenômeno poderia estar diretamente relacionado com uma usual contaminação por rRNAs observada de modo geral nas amostras de mRNAs após purificação por afinidade à oligo(dT). Surpreendentemente, encontramos muitas caudas poli(A)⁺ homo- ou heteropoliméricas na extremidade 3' e internamente localizadas nos transcritos preditos (rRNA e mRNA) interrompendo a informação e provavelmente relacionadas com um mecanismo ainda não observado de controle pós-transcricional da expressão gênica em *Leishmania*. Comprovamos estas observações por RT-PCR e sequenciamento convencional usando os genes ribossomais SSU 18S, LSU 28S alfa e LSU 28S beta e mensageiros eiF5a2 e proteína 60S ribossomal como alvos experimentais. O fenômeno de adição de cauda poli(A)⁺ em moléculas truncadas expande as fronteiras do conhecimento a respeito da regulação da expressão gênica em *Leishmania* e organismos relacionados e sugere a presença de um controle pós-transcricional baseado no "turnover" de RNAs. A presença de transcritos homólogos à proteínas componentes do exossoma de outros organismos foi também observada, sugerindo que esta maquinaria seja ativa em *Leishmania*. A integridade das moléculas transcritas pode ter um papel importante na regulação dos genes que efetivamente são traduzidos. Estes dados destacam a importância de avaliar estas modificações pós-transcricionais nas análises globais do transcriptoma e abrem

novos campos de pesquisa nos estudos globais de expressão gênica diferencial em tripanosomatídeos.

ABSTRACT

SANTOS, Ramon de Freitas, D.Sc., Universidade Federal de Viçosa, June of 2012. **Polyadenylation patterns in RNA molecules of *Leishmania* by transcriptome analysis.** Adviser: Juliana Lopes Rangel Fietto. Co-advisers: Márcia Rogéria de Almeida, Luís Carlos Crocco Afonso and Luciano Gomes Fietto.

The high throughput next generation sequencing (NGS) has revealed new aspects on genomes and transcriptomes that was not observed or poor described using small scale sequencing. Using two species of *Leishmania* as a model in a typical transcriptome analysis and the NGS 454 pirosequencing we found non canonical polyadenilations at rRNAs and mRNAs. *Leishmania*'s RNAs, including rRNAs are transcribed as polycistronical long pre-rRNAs, but usually only mRNAs are processed by poly(A)⁺ polymerases that add canonical 3'-poly(A)⁺ tails. Analyzing the transcriptome data we observed poly(A)⁺ rich tails at the end of rRNA subunits too. This phenomenon could be direct related with the common high rRNA contaminants observed in general mRNA samples after purification by oligo(dT) affinity approach. Surprisingly we found many poly(A)⁺ rich tails at 3'-end at the middle of the predicted transcripts (both rRNA and mRNA) interrupting the message and probably related to a new post-transcriptional control of gene expression in *Leishmania*. We proved these founds by conventional RT-PCR and sangerSanger sequencing using 28S beta rRNA, 18S rRNA, as rRNA targets and 60S ribosomal protein L5 and eiF5a2 mRNA targets. The internal poly(A)⁺-rich tailed RNA truncation phenomenon open new fields on post-transcriptional gene expression regulation in *Leishmania* and related organisms. It adds further complexity to the paradigm that almost the entire trypanosomatids genomes are constitutively expressed, in that the integrity of messages could play an important role in regulating the pool of genes that are effectively translated. These data highlight the importance to evaluate this RNA modification in high throughput transcriptome analyses opening new fields on the studies of transcriptome in general and trypanosomatids gene expression and transcripts stabilities in particular.

INTRODUÇÃO

Nosso grupo de pesquisa estuda há alguns anos a relação entre a virulência dos parasitos do gênero *Leishmania* e de *Trypanosoma cruzi* em relação a diferentes espécies e cepas, focando especialmente a relação entre as ecto-nucleotidases e o controle da infecção ou resposta imune do hospedeiro (Maioli, Takane *et al.*, 2004; De Almeida Marques-Da-Silva, De Oliveira *et al.*, 2008; Santos, Possa *et al.*, 2009; De Souza, De Assis *et al.*, 2010). Atualmente, estamos interessados também em estudar as interações parasito-hospedeiro nas diferentes cepas e espécies de forma mais global, por isto estamos investindo novos esforços em análises transcriptômicas com tecnologias recentemente desenvolvidas de sequenciamento em larga escala de RNAs.

A ideia original se iniciou com a aprovação do presente projeto através do Edital de Pesquisa Interinstitucional (GENOPROT/2007) do Ministério de Ciência e Tecnologia em parceria com a Agência Financiadora de Estudos e Projetos – MCT/FINEP. Nesta parceria foi formada uma rede de grupos de pesquisa, coordenada pela Dra. Maria Júlia Manso Alves do Departamento de Bioquímica do Instituto de Química da USP – São Paulo, com o objetivo de explorar genomas recém sequenciados que tiveram participação de pesquisadores brasileiros. Neste projeto foi adquirida a plataforma de pirosequenciamento (Roche/454 GS FLX Titanium). O projeto geral era denominado “Identificação de Novos Genes e Proteínas de Interesse Biotecnológico para o Brasil”- sigla-DBQ-2007, e incluía o projeto da análise do transcriptoma de *Leishmania (Viannia) braziliensis*. Os resultados apresentados nessa tese de doutorado são partes desta linha de pesquisa em desenvolvimento com o objetivo geral de identificar novos genes alvos para futuros estudos de aplicações biotecnológicas como desenvolvimento de vacinas e desenho racional de drogas.

Para entender melhor o processo de infecção, objetivamos inicialmente explorar as informações genômicas recém-liberadas (Peacock, Seeger *et al.*, 2007) através da análise comparativa do transcriptoma das formas infectantes promastigotas metacíclicas e não infectantes promastigotas procíclicas de um isolado altamente virulento/infectante (isolado NSL) e outro de baixa virulência (isolado ET). Dados internos do nosso grupo de pesquisa e de equipes colaboradoras, produzidos em modelo murino, apresentaram os mesmos perfis de lesão e

características biológicas encontradas em pacientes humanos e comprovaram o comportamento polar entre estes isolados de *L. (V.) braziliensis* quanto à virulência e infectividade (Fietto, J.L.R. & Afonso, L.C.C., comunicação pessoal).

O RNA-Seq (sequenciamento de RNA) usa tecnologias de sequenciamento maciço e de alto rendimento recentemente desenvolvidas para mapear e quantificar os transcriptomas. Em geral, uma população de RNAs (total ou fracionados, tais como os poli(A)⁺ RNAs) é convertida em uma biblioteca de fragmentos de cDNAs com adaptadores ligados em uma ou nas duas extremidades. Cada molécula, com ou sem amplificação, é então sequenciada com alto rendimento e maciçamente paralelizada para obter curtas sequências de uma (“single-end sequencing”) ou de ambas (“paired-end sequencing”) extremidades. Os *reads*, como são chamados estes fragmentos, são moléculas típicas de 30 a 400 pares de bases, dependendo da tecnologia de sequenciamento utilizada. Em princípio, qualquer uma destas tecnologias de próxima geração (NGS) pode ser usada para RNA-Seq, sendo que o Illumina (Wilhelm, Marguerat *et al.*, 2008), o SOLiD (Cloonan, Forrest *et al.*, 2008) e o 454 (Emrich, Barbazuk *et al.*, 2007) são os sistemas mais comumente usados para esta proposta.

Seguinte ao sequenciamento, os *reads* resultantes são então alinhados ao genoma e, ou, transcriptoma de referência, ou são montados *de novo* sem uma sequência genômica para produzir um mapa de transcrição em escala genômica que consiste de uma estrutura transcricional e, ou, do nível de expressão de cada gene. Embora o RNA-Seq seja ainda uma tecnologia que está em constante desenvolvimento, ela oferece atualmente várias vantagens na sua escolha sobre as demais tecnologias existentes. Primeiro, ao contrário de abordagens baseadas em hibridização, o RNA-Seq não está limitado à detecção de transcritos que correspondem às sequências genômicas existentes. Por exemplo, o RNA-Seq baseado na plataforma Illumina foi usado para sequenciar o transcriptoma do patógeno humano *Trypanosoma brucei* e foi capaz de identificar mais de 1000 novos transcritos ainda não catalogados (Kolev, Franklin *et al.*, 2010). De maneira similar, um estudo baseado na plataforma do 454 foi capaz de identificar 10 genes diferencialmente expressos por edição alternativa (“splicing”) em células malignas de mesotelioma pleural, possibilitando aplicações úteis como marcadores

diagnósticos e a identificação de potenciais alvos terapêuticos nestas células tumorais (Dong, Jensen *et al.*, 2009). Além do mais, *reads* curtos de 30 pares de bases derivados de RNA-Seq dão informação a respeito de como dois exons estão conectados, enquanto que *reads* mais longos revelam a conectividade entre múltiplos exons (Mortazavi, Williams *et al.*, 2008; Nagalakshmi, Wang *et al.*, 2008). Estes fatores tornam o RNA-Seq útil para o estudo de transcriptomas mais complexos. Adicionalmente, o RNA-Seq também pode revelar variações na sequência (por exemplo, SNPs) nas regiões transcritas (Barbazuk, Emrich *et al.*, 2007; Cloonan, Forrest *et al.*, 2008; Morin, Bainbridge *et al.*, 2008).

Uma segunda vantagem do RNA-Seq sobre os microarranjos de DNA é que os níveis de “background” são muito baixos ou inexistentes porque não há um limite superior para quantificação, se correlacionando diretamente com o número de sequências obtidas. Consequentemente apresenta um grande intervalo dinâmico dos níveis de expressão sobre os quais os transcritos podem ser detectados (Mortazavi, Williams *et al.*, 2008; Kolev, Franklin *et al.*, 2010; Siegel, Hekstra *et al.*, 2010). Contrastando esta ideia, os microarranjos de DNA não têm sensibilidade para genes expressos em níveis muito baixos ou muito altos, diminuindo muito o intervalo dinâmico de detecção (Okoniewski e Miller, 2006; Royce, Rozowsky *et al.*, 2007). Por fim, como não há passos de clonagem e subclonagem durante a geração das bibliotecas transcriptômicas, o RNA-Seq requer muito menos amostra de RNA e consequentemente custos muito menores do que outros métodos convencionais. Considerando todas estas vantagens em conjunto, esta nova tecnologia é o primeiro método baseado em sequenciamento que permite que o transcriptoma inteiro seja pesquisado de maneira quantitativa e com alto rendimento.

Diante deste contexto e como mencionado anteriormente, no ponto inicial do desenvolvimento do projeto de pesquisa tínhamos o objetivo de estudar as regiões transcritas de *L. (V.) braziliensis* em escala genômica usando a abordagem do RNA-Seq. O passo inicial foi isolar as moléculas maiores de RNAs para serem fragmentadas em pedaços menores (400 a 800 pb), compatível com a metodologia do 454/Roche escolhida. Para isso, usamos a estratégia de purificação com “beads” magnéticos conjugados a oligo(dT) para isolar as moléculas de mRNAs poli(A)⁺ na extremidade 3'. Surpreendentemente, no primeiro passo da construção das

bibliotecas transcriptômicas observamos um persistente alto conteúdo de rRNAs nas amostras poli(A)⁺ purificadas. A descoberta de que os rRNAs em *Leishmania* eram poliadenilados foi corroborada por outros estudos já publicados em diferentes organismos eucarióticos, o que nos motivou a estudar esse fenômeno mais detalhadamente.

A partir de então, mudamos o foco de trabalho e os resultados apresentados nesta tese de doutorado descrevem este evento de poliadenilação em moléculas de RNAs e mais um fenômeno de poliadenilação ainda não observado em *Leishmania*, a poliadenilação interna truncando genes neste importante patógeno humano. Assim, neste trabalho, mostramos que os rRNAs são poliadenilados na extremidade 3' em *L. (V.) braziliensis* e *L. (L.) amazonensis*, similar ao que ocorre nos mRNAs de eucariotos. Além disto, analisamos a presença de genes codificantes de proteínas do exossoma, principal complexo protéico regulador do “turnover” de RNAs, no genoma de *L. (V.) braziliensis* e verificamos que todas as proteínas do complexo são ativamente transcritas nesta espécie. Isso abre uma nova fronteira para estudo de regulação da expressão gênica em *Leishmania* podendo também, de um modo mais geral, ser aplicada para tripanossomatídeos.

REVISÃO BIBLIOGRÁFICA

***Leishmania* e Leishmanioses**

As espécies de *Leishmania* são protozoários da Ordem Kinetoplastida. Todas são parasitos intracelulares obrigatórios que tem dois estágios evolutivos no seu ciclo de vida. As formas alongadas vivem e replicam no intestino de fêmeas de flebotomíneos (insetos vetores do gênero *Lutzomya* ou *Phlebotomus*, dependendo da localização geográfica) como formas promastigotas flageladas. Uma variação das formas promastigotas são as formas promastigotas metacíclicas, localizadas na porção anterior do aparelho digestivo das fêmeas de flebotomíneos e responsáveis pela infecção dos hospedeiros vertebrados através do mecanismo natural de infecção (durante o repasto sanguíneo). Estes promastigotas metacíclicos invadem ou são fagocitados por células fagocíticas e imediatamente recrutam outras células hospedeiras.

As formas amastigotas se replicam dentro de fagolisossomos no interior de células fagocíticas mononucleares de mamíferos, primariamente macrófagos e células dendríticas, frequentemente induzindo respostas inflamatórias e persistindo no tecido infectado. Estes amastigotas ainda são capazes de infectar outros macrófagos vizinhos ou macrófagos em regiões mais distantes, culminando no desenvolvimento de diferentes formas clínicas (Murray, Berman *et al.*; 2005). Porém, é bem documentado que em lesões na pele, outros tipos de células, incluindo neutrófilos e fibroblastos, também são infectadas (Bogdan, Donhauser *et al.*, 2000; Laskay, Van Zandbergen *et al.*, 2003). Embora a variabilidade genética do hospedeiro e as respostas imunes específicas, juntamente com as espécies de insetos vetores transmissores e fatores ambientais, sejam bem conhecidos por influenciar o resultado das infecções, o principal fator que determina as apresentações clínicas são as espécies infectantes do parasito (Lipoldova e Demant, 2006).

Por exemplo, o parasito do Novo Mundo *L. (V.) braziliensis* é o agente causal de leishmaniose mucocutânea, enquanto que as espécies do Velho Mundo *L. (L.) major* e *L. (L.) infantum*, presentes na África, Europa e Ásia, são parasitos que causam leishmaniose cutânea e visceral, respectivamente. Outro exemplo determinante do resultado da infecção espécie-dependente é o desenvolvimento de vacúolos parasitóforos (VPs) morfologicamente distintos. Enquanto parasitos do complexo *L. mexicana* [*L. (L.) mexicana*, *L. (L.) amazonensis* e *L. (L.) pifanoi*]

residem em VPs comuns que se tornam progressivamente distendidos, parasitos do complexo *L. donovani* [*L. (L.) donovani*, *L. (L.) infantum* e *L. (L.) chagasi*], na maioria das vezes, habitam VPs apertados e individuais onde as células filhas segregam em seus próprios VPs (Kima, 2007). Tais diferenças no resultado das interações entre estes parasitos e macrófagos são uma poderosa informação de que as espécies de *Leishmania* são organismos diversos com significativas diferenças em nível genético e biológico, que resultam em diferenças em múltiplos pontos em suas interações com as células hospedeiras.

As leishmanioses têm diversas manifestações clínicas: lesões ulcerativas na pele, inflamação destrutiva da mucosa e infecção visceral (Kala-azar) disseminada. A epidemiologia, imunopatologia e resultado são similarmente diversos, uma vez que as infecções ocorrem em múltiplas regiões endêmicas, tanto em crianças quanto em adultos. Ao menos vinte espécies de *Leishmania* infectam seres humanos e são transmitidas por aproximadamente trinta espécies de flebotomíneos (Pearson e Sousa, 1996; Herwaldt, 1999; Desjeux, 2001; 2004). A variação da intensidade de desenvolvimento ou manifestação da doença também tem sido demonstrada em animais naturalmente infectados e, especialmente, animais experimentalmente infectados (Alvar, Canavate *et al.*, 2004; McMahon-Pratt e Alexander, 2004).

Não obstante, todas as formas de infecção por este protozoário compartilham três características patogênicas: (i) macrófagos nos tecidos infectados como células alvos suportando a replicação intracelular do parasito; (ii) respostas imunoinflamatórias do hospedeiro regulando a expressão e o resultado da doença; e (iii) a persistência residual da infecção nos tecidos de tropismo correspondentes a cada espécie de *Leishmania* (Murray, Berman *et al.*, 2005).

As leishmanioses são consideradas primariamente uma zoonose podendo acometer o homem quando este entra em contato com o ciclo de transmissão do parasito, transformando-se em uma antropozoonose. Elas são endêmicas em grandes áreas dos trópicos, subtropicais e na bacia do Mediterrâneo, matando milhares de pessoas e desabilitando outras tantas a cada ano (Desjeux, 2004). Atualmente, encontra-se entre as seis endemias consideradas prioritárias no mundo (Yamey e Torreele, 2002). Há cerca de 12 milhões de casos da doença no mundo, dos quais um número estimado de 2 milhões de indivíduos desenvolvem doenças sintomáticas (1 a

1,5 milhão de casos de leishmaniose cutânea; 0,5 milhão de casos de leishmaniose visceral). A incidência de infecção é ainda superior quando casos subclínicos são incluídos (Desjeux, 2004). As leishmanioses estão associadas com cerca de 2 a 4 milhões de indivíduos com algum real prejuízo de saúde e ao redor de 70.000 mortes anuais (Desjeux, 2004). Além disto, por ser endêmica em 88 países, calcula-se que existam hoje cerca de 367 milhões de pessoas em risco de contraí-la. Apesar de serem consideradas doenças rurais, o recente aparecimento, especialmente da forma visceral, em grandes centros urbanos, faz das leishmanioses um problema de saúde pública emergente.

O espectro de doenças que elas causam pode ser categorizado amplamente em três tipos: (i) leishmaniose visceral, a forma mais séria na qual os parasitos saem do sítio de inoculação e proliferam no fígado, baço e medula óssea, resultando em imunossupressão do hospedeiro e morte na ausência do tratamento; (ii) leishmaniose cutânea, na qual os parasitos permanecem no sítio de infecção e causam uma ulceração de longa duração; e (iii) leishmaniose mucocutânea, com destruição crônica de mucosas (Marsden, 1986).

Vários são os reservatórios silvestres com importância relativa para a manutenção do ciclo de vida das várias espécies do protozoário. Algumas espécies de *Leishmania* têm seus reservatórios restritos à família Canidae, sendo o cão doméstico extremamente importante em termos epidemiológicos, tanto no ambiente urbano quanto no rural, devido à alta prevalência da Leishmaniose Canina e ao intenso parasitismo visceral e cutâneo que ocorre nestes animais (Deane e Deane, 1954; Dantas-Torres, 2007; Baneth, Koutinas *et al.*, 2008; Miro, Cardoso *et al.*, 2008).

Cerca de 90% dos casos de leishmaniose cutânea ocorrem no Afeganistão, Paquistão, Síria, Arábia Saudita, Algéria, Irã, Brasil e Peru; 90% dos casos de leishmaniose visceral correm na Índia, Bangladesh, Nepal, Sudão, Etiópia e Brasil (Desjeux, 2004). Em vista desse quadro geográfico de distribuição, as leishmanioses permanecem associadas à pobreza como doenças negligenciadas (Yamey e Torreele, 2002). Com exceção do sul europeu, os órgãos básicos de saúde pública em regiões endêmicas bloqueiam o acesso ao diagnóstico, tratamento e controle efetivo da doença. Com poucas perspectivas de retorno financeiro, o desenvolvimento de drogas efetivas permanece paralisado (Guerin, Olliaro *et al.*, 2002). Contudo, e

apesar destes obstáculos, nos últimos anos a ciência alcançou considerável progresso no diagnóstico, tratamento e controle do vetor (Sundar, Sahu *et al.*, 2002; Davies, Kaye *et al.*, 2003; Vega-Lopez, 2003; Murray, 2004; Soto, Arana *et al.*, 2004). Pesquisas nesta área tem aumentado a compreensão da susceptibilidade, dos mecanismos que regulam o desenvolvimento da doença (Blackwell, Mohamed *et al.*, 2004) e as múltiplas interações parasito-hospedeiro; a imunidade contra estes parasitos tem sido mais precisamente decifrada e vacinas estão em desenvolvimento (Requena, Iborra *et al.*, 2004; Coler e Reed, 2005).

A sequência dos genomas de algumas espécies de *Leishmania* e dos vetores flebotomíneos (Ivens, Peacock *et al.*, 2005; Peacock, Seeger *et al.*, 2007) estão a alguns anos disponíveis e estão facilitando os avanços neste campo da ciência para solucionar questões fundamentais como a biologia do parasito, mecanismos de transmissão e patogênese, fornecendo assim, novos genes alvos para ensaios de diagnósticos, desenvolvimento de drogas e, ou, vacinas.

Espectro diverso das leishmanioses

As propriedades do parasito (infectividade, patogenicidade e virulência) (Weigle e Saravia, 1996), os fatores inerentes ao hospedeiro e as suas respostas, regulam a expressão heterogênea das doenças e suas respectivas manifestações clínicas; a expressão e as manifestações variam, ainda, em função das espécies do parasito e da região endêmica. Quatro exemplos ilustram algumas destas diversidades básicas entre as leishmanioses.

Infecção subclínica

As infecções causadas por *Leishmania*, especialmente as oriundas das espécies viscerotrópicas, podem permanecer completamente assintomáticas (Weigle e Saravia, 1996; Le Fichoux, Quaranta *et al.*, 1999; Follador, Araujo *et al.*, 2002; Guerin, Oliaro *et al.*, 2002; Murray, 2004; Riera, Fisa *et al.*, 2004). Esta observação chama atenção para fatores do hospedeiro na suscetibilidade e expressão da doença – idade, estado nutricional e eficácia das respostas imunes inatas e adquiridas (Cerf, Jones *et al.*, 1987; Carvalho, Barral *et al.*, 1992; Louzir, Melby *et al.*, 1998; Davies e Mazloumi Gavvani, 1999; Davies, Reithinger *et al.*, 2000; Machado-Coelho, Caiaffa *et al.*, 2005). Este último fator, por sua vez, pode ser afetado por

polimorfismos imunogenéticos (Karplus, Jeronimo *et al.*, 2002; Blackwell, Mohamed *et al.*, 2004).

Infecção localizada versus disseminada

De modo geral, as infecções mais clinicamente aparentes causadas por *Leishmania*, permanecem localizadas na pele ou nos nódulos linfáticos adjacentes. Não obstante, certas espécies escapam para a mucosa nasal e, ou, orofaríngea, sítios cutâneos múltiplos, ou para o fígado, baço, medula óssea e nódulos linfáticos distantes. Estas disseminações apontam principalmente, mas não exclusivamente, para diferentes propriedades inerentes aos parasitos – sensibilidade à temperatura (Reithinger, Mohsen *et al.*, 2005), tropismo tecidual (Zhang, Mendez *et al.*, 2003) e capacidade para imunoevasão e persistência (Mcdowell, Marovich *et al.*, 2002). Como estes parasitos alcançam a circulação sanguínea sistêmica ainda não é compreendido. Os amastigotas circulam ativamente e a parasitemia é frequente nos pacientes HIV-imunodeficientes (Pintado, Martin-Rabadan *et al.*, 2001). Paradoxalmente, entretanto, a parasitemia assintomática também ocorre em indivíduos imunodeficientes com quadro de infecção visceral subclínica (Le Fichoux, Quaranta *et al.*, 1999; Riera, Fisa *et al.*, 2004).

Resposta do hospedeiro à infecção

As respostas das células T e a ativação de macrófagos induzidos por citocinas, determinantes da imunidade mediada por células, a expressão ou desenvolvimento da doença, e a carga parasitária, são também particularmente variáveis. A hipersensibilidade atrasada (teste dérmico positivo), reatividade das células T antígeno-específico, e a secreção de citocinas ativadoras são, por exemplo, mantidas em infecções sintomáticas e em leishmanioses cutâneas localizadas (Meller-Melloul, Farnarier *et al.*, 1991; Carvalho, Barral *et al.*, 1992; Weigle e Saravia, 1996; D'oliveira Junior, Costa *et al.*, 1997; Louzir, Melby *et al.*, 1998; Ajdary, Alimohammadian *et al.*, 2000; Gama, Costa *et al.*, 2004); estão presentes, mas ineficazes, para leishmaniose visceral e leishmaniose dérmica pós Kala-azar (Sundar, Reed *et al.*, 1997; Zijlstra, Musa *et al.*, 2003; Gama, Costa *et al.*, 2004; Hailu, Van DerPoll *et al.*, 2004); e exuberante, mas aparentemente patológica, em infecções nas mucosas (Cabrera, Shaw *et al.*, 1995; Bacellar, Lessa *et al.*, 2002).

Variações regionais

As características clínicas básicas, tanto da leishmaniose cutânea quanto da leishmaniose visceral, também variam substancialmente entre regiões endêmicas, e também entre elas, provavelmente indicando interações entre as propriedades locais dos parasitos, a biologia dos vetores e os fatores dos hospedeiros. As variações regionais incluem: população primariamente direcionada (crianças jovens versus adolescentes e adultos), modo de transmissão (zoonótica versus antroponótica), descobertas clínicas particulares e síndromes, probabilidade e ritmo de cura espontânea, e ritmo de recorrência. As abordagens de tratamento e respostas à quimioterapia também variam em função da região. Assim, a diversidade das leishmanioses, cutânea e visceral, frequentemente torna necessária a subclassificação clínica endêmica regional (Desjeux, 2001; 2004).

Similaridades básicas

Felizmente, características compartilhadas ajudam no entendimento da patogênese, defesas do hospedeiro, problemas de persistência crônica e reativação, na elaboração de novas abordagens para o tratamento e, esperançosamente, prevenção por vacinação. Nas leishmanioses a porta de entrada é a mesma, os macrófagos teciduais são as células alvos, a inflamação e a imunidade ativada por macrófagos regulam a expressão da doença e o prognóstico inicial, e a persistência do parasito é a responsável por estas ações, apesar da cura espontânea ou da resolução clínica induzida por tratamento (Le Fichoux, Quaranta *et al.*, 1999; Belkaid, Hoffmann *et al.*, 2001; Sacks e Noben-Trauth, 2002; Dereure, Duong Thanh *et al.*, 2003; Engwerda, Ato *et al.*, 2004; Mendonca, De Brito *et al.*, 2004; Murray, 2004; Riera, Fisa *et al.*, 2004).

Leishmaniose cutânea e mucocutânea

A Leishmaniose Tegumentar Americana (LTA) é primariamente uma infecção zoonótica, afetando outros animais que não o ser humano, o qual pode ser envolvido secundariamente. Esta doença é caracterizada tanto por sua diversidade quanto pela complexidade (Herwaldt, 1999). Uma pápula tipicamente começa na picada do flebotômíneo, aumenta para um nódulo, e ulcera após 1 a 3 meses (Dowlati, 1996; Machado, Araujo *et al.*, 2002; Magill, 2005). Placas planas e

hiperqueratósicas ou “verruca-like” também se desenvolvem na doença do velho mundo. Pacientes, incluindo viajantes e militares (Blum, Desjeux *et al.*, 2004; Weina, Neafie *et al.*, 2004; Magill, 2005), primeiramente chamam atenção porque desenvolvem de uma a duas, ou às vezes até várias (até dezenas) de lesões na pele.

Na leishmaniose do velho mundo, a maioria das lesões são papulares, nodulares, ou nódulo-ulcerativas (Dowlati, 1996), enquanto que lesões ulcerativas são mais comuns nas lesões cutâneas da leishmaniose do novo mundo (Palacios, Osorio *et al.*, 2001). As lesões disseminadas (Turetz, Machado *et al.*, 2002) e linfadenopatias localizadas precedendo as úlceras na pele (Barral, Guerreiro *et al.*, 1995) ocorrem principalmente no Brasil. A espécie *L. (V.) braziliensis* foi a primeira espécie de *Leishmania* descrita e confirmada como agente etiológico da LTA. É a mais importante, não só no Brasil, mas em toda a América Latina. A disseminação mucosa ocorre em 1 a 10% das infecções, se desenvolvendo 1 a 5 anos após a lesão cutânea ter se curado, mas às vezes é coincidente com lesões cutâneas ativas; cerca de 90% dos pacientes tem uma cicatriz cutânea precedente (Weigle e Saravia, 1996; Madeira Md Mde, Uchoa *et al.*, 2003; Blum, Desjeux *et al.*, 2004; Machado-Coelho, Caiaffa *et al.*, 2005; Madeira Md Mde, Schubach Ade *et al.*, 2005; Madeira, Schubach *et al.*, 2006; Dantas-Torres, 2007).

As leishmanioses mucosas começam com um eritema e ulcerações nas narinas, procedente da perfuração do septo nasal e de lesões inflamatórias destrutivas. Estas últimas podem obstruir a faringe ou a laringe e produz considerável desfiguração (Franke, Wignall *et al.*, 1990). A doença mucosa é ocasionalmente relatada fora da América Latina (Osman, Kager *et al.*, 2000; Aliaga, Cobo *et al.*, 2003) e pode ser adquirida por viajantes (Ahluwalia, Lawn *et al.*, 2004).

Na leishmaniose cutânea difusa, raramente vista na Etiópia e na América Latina, os nódulos carregados de parasitos são generalizados e não ulcerados (Bryceson, 1969; Convit, Ulrich *et al.*, 1993). Na leishmaniose pós Kala-azar, uma doença de patogênese incerta que ocorre no leste da África e no subcontinente Indiano [*L. (L.) donovani*], uma erupção nodular maculopapular é muitas vezes primeiramente localizada ao redor da boca e então se torna generalizada (Zijlstra, Musa *et al.*, 2003). No Sudão esta forma da doença se manifesta em cerca da metade dos pacientes dentro de aproximadamente 6 meses após o diagnóstico da

leishmaniose visceral, mas se cura espontaneamente dentro de 6 meses e usualmente não precisa de tratamento adicional (Zijlstra, Musa *et al.*, 2003). Na Índia a leishmaniose pós Kala-zar se desenvolve em cerca de 5 a 10% dos pacientes vários anos após o tratamento aparentemente bem sucedido para leishmaniose e precisa novamente de um tratamento (Zijlstra, Musa *et al.*, 2003).

A doença cutânea cura por reconstituição do epitélio e posterior cicatrização na região da lesão. A maioria das lesões no velho mundo cura espontaneamente dentro de 2 a 4 meses [*L. (L.) major*] (Asilian, Jalayer *et al.*, 1995; Ben Salah, Zakraoui *et al.*, 1995; Dowlati, 1996; Alrajhi, Ibrahim *et al.*, 2002) ou de 6 a 15 meses para *L. (L.) tropica* (Dowlati, 1996; Blum, Desjeux *et al.*, 2004). Para as leishmanioses cutâneas do novo mundo, a cura espontânea ocorre após 3 meses para *L. (L.) mexicana* (em mais de 75% dos casos) (Herwaldt, Arana *et al.*, 1992), mas ocorre de maneira mais lenta em *L. (V.) braziliensis* (cerca de 10% dos casos) e também nas infecções causadas por *L. (V.) panamensis* (aproximadamente 35% dos casos) (Soto-Mancipe, Grogl *et al.*, 1993; Velez, Agudelo *et al.*, 1997; Soto, Arana *et al.*, 2004).

O tratamento da leishmaniose cutânea é realizado para acelerar a cura, reduzindo as cicatrizes no local afetado e prevenindo sua disseminação para outras regiões (como, por exemplo, para as mucosas) ou recaída. O tratamento é especialmente indicado para pacientes com lesões persistentes (superior a seis meses) ou lesões que estão localizadas sobre articulação, múltiplas (5 a 10 ou mais), ou grandes (4 a 5 cm ou superior) (Blum, Desjeux *et al.*, 2004; Weina, Neafie *et al.*, 2004). O antimônio parenteral apresenta uma eficiência limitada na cura da leishmaniose cutânea em todas as regiões. Além do mais, o antimônio é caro, exceto em formulações genéricas (Sampaio, Castro *et al.*, 1971; Jha, 1983; Arana, Navin *et al.*, 1994; Alkhawajah, Larbi *et al.*, 1997; Aronson, Wortmann *et al.*, 1998; Jha, Olliaro *et al.*, 1998; Romero, Vinitius De Farias Guerra *et al.*, 2001; Salmanpour, Handjani *et al.*, 2001; Sundar, Jha *et al.*, 2002; Asilian, Sadeghinia *et al.*, 2003; Soto, Valda-Rodriguez *et al.*, 2004; Sundar, Mehta *et al.*, 2004; Uzun, Durdu *et al.*, 2004; Andersen, Cruz-Saldarriaga *et al.*, 2005).

Ainda há a necessidade do desenvolvimento de agentes orais mais baratos e seguros, cuidadosamente testados e com boa tolerância. Considerando que não há

consenso sobre o tratamento ideal para as leishmanioses cutâneas em geral, alternativas ao antimônio estão sob investigação ativa. Com exceção do antimônio inoculado diretamente dentro da lesão, o qual é usado de uma maneira altamente variável, outros esquemas alternativos têm sido quase sempre avaliados em estudos individuais. De um modo geral, vacinas não estão disponíveis e as drogas em uso são altamente tóxicas, têm regimes terapêuticos complexos e há relatos de indução de resistência em algumas regiões endêmicas (Croft e Yardley, 2002; Desjeux, 2004; Gourbal, Sonuc *et al.*, 2004).

Leishmaniose visceral

A leishmaniose visceral (LV) é uma doença sistêmica que é fatal se não tratada e é causada por espécies do complexo *Leishmania donovani* [*L. (L.) donovani*, *L. (L.) infantum*, *L. (L.) chagasi*] no leste da África, no subcontinente Indiano, na Europa, norte da África e na América Latina (Mauricio, Stothard *et al.*, 2000; Lukes, Mauricio *et al.*, 2007; Smith, Peacock *et al.*, 2007). Há dois tipos de LV, nas quais diferem nas características de transmissão: a LV zoonótica é transmitida do animal para o vetor e depois para o homem; a LV antroponótica é transmitida do humano para o vetor e para o homem novamente. Os humanos são os hospedeiros ocasionais e os animais, principalmente os cães, são os reservatórios do parasito (Alvar, Canavate *et al.*, 2004).

São estimados 500.000 novos casos de LV e mais de 50.000 mortes ocasionadas pela doença a cada ano (Desjeux, 2004), um número de mortes que é superado entre as doenças parasitárias apenas pela malária (www.who.int/whr/2002/en/whr02_en.pdf). Ambos os valores são aproximações uma vez que a LV muitas vezes não é reconhecida ou não relatada (Collin, Coleman *et al.*, 2006; Singh, Reddy *et al.*, 2006). A maioria (mais de 90%) dos casos ocorre em apenas seis países: Bangladesh, Índia, Nepal, Sudão, Etiópia e Brasil. A migração, a falta de medidas de controle e a co-infecção LV-HIV são os três principais fatores que impulsionam o aumento da incidência de LV (Boelaert, Criel *et al.*, 2000; Desjeux, 2001). Graves epidemias de LV foram relatadas no passado: no sul do Sudão, em um contexto de guerra civil e fome, a LV matou cerca de 100.000 pessoas de uma população de 280 mil entre 1984 e 1994 (Seaman, Mercer *et al.*, 1996).

A LV afeta comunidades pobres, geralmente em áreas rurais remotas. A doença é endêmica na maior parte dos países que estão entre os menos desenvolvidos do mundo (como o Nepal) ou em regiões mais pobres dos chamados países “de renda média” (como o estado de Bihar, na Índia). Pacientes e famílias afetadas por LV tornam-se ainda mais pobres por causa dos elevados custos diretos (por exemplo, por causa dos custos de diagnóstico e tratamento da LV) e dos custos indiretos (por exemplo, perda de renda familiar) por causa da doença (Ahluwalia, Bern *et al.*, 2003; Alvar, Yactayo *et al.*, 2006; Rijal, Koirala *et al.*, 2006). Como Índia, Nepal e Bangladesh albergam 67% da LV estimada no mundo (Hotez, Remme *et al.*, 2004), o compromisso dos governos desses países para lançar um programa regional de eliminação da LV é sempre bem vindo. O objetivo deste programa é eliminar a LV como um problema de saúde pública desses países até 2015, através de uma abordagem local para reduzir a incidência anual de LV para menos de um caso por 10.000 indivíduos (Chappuis, Sundar *et al.*, 2007).

Após um período de incubação que dura geralmente entre 2 a 6 meses, os pacientes com LV apresentam sinais e sintomas de uma infecção sistêmica (incluindo febre persistente, fadiga, fraqueza, perda de apetite e perda de peso) e invasão parasitária no sangue e no sistema retículo-endotelial (isto é, o sistema geral fagocítico), tais como aumento dos gânglios linfáticos, baço e fígado. A febre é geralmente associada com calafrios e pode ser intermitente. Fadiga e fraqueza são agravadas por anemia, que é causada pelo estado inflamatório persistente, hiperesplenismo (a destruição periférica dos eritrócitos no baço) e, por vezes, por sangramento (Desjeux, 2004; Chappuis, Sundar *et al.*, 2007).

A apresentação clínica da LV é semelhante nas diversas áreas endêmicas, mas existem algumas diferenças. Por exemplo, aumento dos gânglios linfáticos é raramente encontrados em pacientes indianos com LV, mas são frequentes em pacientes com LV no Sudão (Siddig, Ghalib *et al.*, 1990; Zijlstra, Ali *et al.*, 1991). A hiperpigmentação, o que provavelmente deu origem ao nome kala-azar (febre negra em Hindu), só foi descrita em pacientes com LV no subcontinente Indiano, mas hoje este sintoma é incomum e foi, talvez, uma característica de doença prolongada na época em que o tratamento eficaz não estava disponível. Com o avanço da doença, a esplenomegalia pode aumentar, causando distensão e dores abdominais, que é por vezes aumentada por concomitante hepatomegalia. Sintomas e sinais de co-infecções

bacterianas, tais como diarreia, pneumonia ou tuberculose pode confundir o quadro clínico no momento do diagnóstico inicial. Os sintomas da LV frequentemente persistem por várias semanas a meses antes de os pacientes ou procurarem assistência médica ou morrerem de co-infecções bacterianas, sangramento forte ou anemia grave (Chappuis, Sundar *et al.*, 2007).

Os parasitos intracelulares replicantes (formas amastigotas) se difundem através do sistema linfático e vascular e infectam outras células fagocíticas no sistema retículo-endotelial, o que resulta em infiltração de medula óssea, hepatoesplenomegalia e, por vezes, gânglios linfáticos (linfadenopatia). É importante ressaltar que a infecção nem sempre equivale com a doença clínica. A proporção de infecções assintomáticas para casos clínicos varia, sendo 1:2,6 a 11:1 no Sudão (Zijlstra, El-Hassan *et al.*, 1994; Khalil, Zijlstra *et al.*, 2002), 4:1 no Kênia (Schaefer, Kurtzhals *et al.*, 1995), 5,6:1 na Etiópia (Ali e Ashford, 1994b; a), 13:1 no Irã (Davies e Mazloumi Gavvani, 1999), 8:1 a 18:1 no Brasil (Badaro, Jones *et al.*, 1986; Evans, Teixeira *et al.*, 1992) e 50:1 na Espanha (Moral, Rubio *et al.*, 2002).

É de extrema importância o desenvolvimento de vacinas e controle da doença para entender os fatores que podem predispor algumas pessoas a desenvolver a doença ou a capacidade de controlar a infecção. A resposta imune específica do hospedeiro mediada por células tem um papel importante no controle da infecção. Em pacientes com LV, a incapacidade de controlar a infecção por *L. (L.) donovani* está associada a uma profunda falta de responsividade das células T aos antígenos de *L. (L.) donovani* (Sacks, Lal *et al.*, 1987) e à produção de interleucina 10 (IL10) (Ghalib, Piuvezamet *et al.*, 1993). As células T CD25-Foxp3-produtoras de IL10 foram recentemente implicada na patogênese da LV humana na Índia (Nylen, Maurya *et al.*, 2007). O papel crucial da resposta imune específica do hospedeiro mediada por células é ilustrado pelo aumento do risco de desenvolvimento de doença clínica em casos de má nutrição ou concomitantes doenças imunossupressoras, tais como a infecção por HIV (Cerf, Jones *et al.*, 1987; Alvar, Canavate *et al.*, 1997; Murray, 1999). Outros fatores de risco para o desenvolvimento de doença clínica foram identificados e incluem a idade jovem (Evans, Teixeira *et al.*, 1992; Zijlstra, El-Hassan *et al.*, 1994; Davies e Mazloumi Gavvani, 1999), diminuição da produção de interferon γ (IFN- γ) (Carvalho, Barral *et al.*, 1992) e polimorfismos no promotor do gene do fator de necrose tumoral α (TNF- α) (Karplus, Jeronimo *et al.*, 2002).

Tendo em vista o número expressivo de casos da doença no mundo, bem como no Brasil, a morbidade causada pelas diferentes formas clínicas da doença e a limitação dos tratamentos e profilaxia para humanos e animais, especialmente cães, com as drogas atualmente disponíveis, é de extrema importância o desenvolvimento de novas abordagens de diagnóstico, tratamento e prevenção. Neste contexto, a descoberta de genes reguladores de virulência do processo de infecção presentes nas interações parasito-hospedeiro, o conhecimento da função destas biomoléculas e o entendimento geral dos mecanismos de regulação da expressão gênica, têm grande potencial para ser usado em abordagens como desenho racional de novas drogas, vacinas ou desenvolvimento de métodos de diagnóstico.

Arquitetura e organização genômica em *Leishmania* spp.

Os tripanosomatídeos ocupam uma posição muito divergente dentro do ramo eucariótico da vida, e se separaram da linhagem metazoária três bilhões de anos atrás (Momen e Cupolillo, 2000; Kerr, 2006). O recente término dos sequenciamentos genômicos de muitos protozoários patogênicos humanos, incluindo os genomas das três principais espécies modelos experimentais da ordem Kinetoplastida – *T. brucei*, *T. cruzi* e *L. (L.) major* – ofereceu uma oportunidade para investigar os mecanismos que regulam o desenvolvimento em nível molecular e utilizar esta informação numa nova era pós-genômica (Berriman, Ghedin *et al.*, 2005; El-Sayed, Myler, Blandin *et al.*, 2005; Ivens, Peacock *et al.*, 2005). Muitos aspectos destes organismos são extremamente incomuns, apresentando uma série notável de processos biológicos ímpares entre os eucariotos, particularmente no que se refere à biogênese e metabolismo do RNA. No genoma destes protozoários, os genes estão organizados nos cromossomos como grandes clusters direcionais (do inglês, DGC) que são constitutivamente co-transcritos pela RNA Polimerase II (Pol II) produzindo pré-mRNAs policistrônicos que são sequencialmente processados em mRNAs individuais por “*trans* splicing” e poliadenilação (Lebowitz, Smith *et al.*, 1993; Liang, Haritan *et al.*, 2003; El-Sayed, Myler, Blandin *et al.*, 2005; Peacock, Seeger *et al.*, 2007).

Estes parasitos digenéticos unicelulares alternam seu ciclo de vida entre insetos vetores e hospedeiros mamíferos durante seu desenvolvimento celular. Estas contínuas mudanças ambientais submetem estes organismos a mudanças dinâmicas

na morfologia celular, metabolismo, sinalização celular e na regulação da expressão de conjuntos de genes para permitir uma rápida adaptação às novas condições ambientais (Gilinger e Bellofatto, 2001; Tyler e Engman, 2001; Fenn e Matthews, 2007). A expressão gênica, com exceção dos genes “Spliced Leader RNA”, não depende de promotores canônicos, e assim, não há evidências para o controle da iniciação da transcrição gênica pela modulação da atividade da Pol II (Gilinger e Bellofatto, 2001; Clayton, 2002; Campbell, Thomas *et al.*, 2003; Clayton e Shapira, 2007; Haile, Cristodero *et al.*, 2007). Em contraste ao que ocorre para operons de procariotos, entretanto, as unidades policistrônicas são co-transcricionalmente processadas antes da tradução e não há evidências para o agrupamento funcional dentro destas unidades de transcrição policistrônica (Liang, Haritan *et al.*, 2003). Durante a reação de “*trans* splicing”, um RNA de 39 nucleotídeos (Spliced Leader RNA, SL) contendo um *cap*trimetil é adicionado na extremidade 5’ do mRNA. O sítio acceptor é um dinucleotídeo AG imediatamente após um extensão de polipirimidina (Liang, Haritan *et al.*, 2003; Benz, Nilsson *et al.*, 2005). Os sítios de poliadenilação estão localizados 100 a 300 nucleotídeos acima do sinal de “*trans* splicing” do próximo gene no policistron e estas duas reações de processamento de RNA estão acopladas espacial e temporalmente (Lebowitz, Smith *et al.*, 1993; Clayton e Shapira, 2007; Haile, Cristodero *et al.*, 2007).

Como já mencionado, pouco é conhecido sobre os mecanismos de iniciação da transcrição e somente uns poucos promotores foram analisados (Gilinger e Bellofatto, 2001; Clayton, 2002; Clayton e Shapira, 2007). Evidências experimentais sugerem que a transcrição policistrônica dos DGCs inicia-se e ocorre bidirecionalmente nas “regiões de troca de fitas” (SSR) divergentes e termina nas SSR convergentes que, frequentemente, contêm genes de tRNA, rRNA e, ou, snRNA (El-Sayed, Ghedin *et al.*, 2003; Hall, Berriman *et al.*, 2003; Martinez-Calvillo, Yan *et al.*, 2003; Worthey, Martinez-Calvillo *et al.*, 2003; Martinez-Calvillo, Nguyen *et al.*, 2004).

Os genomas haplóides de *T. brucei*, *L. (L.) major* e *T. cruzi* contêm entre 25 e 55 megabases (Mb) distribuídos em 11 a 36 cromossomos diplóides, e possuem mapeados cerca de 8100, 8300 e 12000 genes codificantes para proteínas, respectivamente (El-Sayed, Myler, Bartholomeu *et al.*, 2005). As espécies de

Leishmania do Velho Mundo (complexos *L. (L.) donovani* e *L. (L.) major*) possuem 36 pares de cromossomos, com 0,28 a 2,8 Mb, enquanto que as espécies do Novo Mundo têm 34 ou 35, com os cromossomos 8+29 e 20+36 fusionados no complexo *L. (L.) mexicana* e 20+34 fusionados na espécie *L. (V.) braziliensis* (Wincker, Ravel *et al.*, 1996; Britto, Ravel *et al.*, 1998).

Uma comparação do conteúdo gênico e da arquitetura do genoma através de um alinhamento local de “all-versus-all” utilizando a ferramenta “BlastP” revelou um centro proteômico conservado de 6158 genes ortólogos em “clusters” policistrônicos (COGs), nos quais foram definidos como “TriTryp core proteome”. Curiosamente, quase todos os COGs (94%) que formam esse centro proteômico estão dentro de regiões de sintenia gênica conservada (El-Sayed, Myler, Blandin *et al.*, 2005). Um alinhamento de sequências de aminoácidos dos produtos gênicos dos COGs desses três tripanosomatídeos revelou uma cobertura de 57% de identidade entre *T. brucei* e *T. cruzi*, e 44% de identidade entre *L. (L.) major* e os outros dois tripanosomas, refletindo os relacionamentos genéticos esperados (Lukes, Jirku *et al.*, 1997; Haag, O’huigin *et al.*, 1998; Stevens, Noyes *et al.*, 1999; Wright, Li *et al.*, 1999). *L. (L.) major* e *T. cruzi* parecem compartilhar ligeiramente mais COGs do que *T. brucei* e *T. cruzi* e consideravelmente mais COGs do que *L. (L.) major* e *T. brucei*. O restante de cada proteoma é composto de membros espécie-específicos, especialmente grandes famílias de antígenos de superfície, dos quais *T. cruzi* (32%) e *T. brucei* (26%) tem uma maior proporção do que *L. (L.) major* (12%). As outras proteínas espécie-específicas realizam distintas funções metabólicas e fisiológicas (El-Sayed, Myler, Blandin *et al.*, 2005).

Expressão gênica diferencial no gênero *Leishmania*

Com o término dos projetos genomas de *L. (L.) major*, *L. (L.) infantum*, *L. (V.) braziliensis*, e de *L. (L.) mexicana* em andamento, as análises quantitativas da expressão gênica global podem ser realizadas nos diferentes estágios de vida das diferentes espécies de *Leishmania*, utilizando diferentes abordagens, com o objetivo de se entender a diferenciação e os passos envolvidos no desenvolvimento desses parasitos.

Até o ano de 2007, as análises de expressão gênica diferencial por microarranjos eram baseadas no genoma de *L. (L.) major* devido ao alto grau de sintenia e

similaridade de sequência com o genoma de outras espécies de *Leishmania* (Peacock, Seeger et al., 2007). Holzer e colaboradores (2006), utilizando sondas oligonucleotídicas de 60-mer, usaram microarranjos para analisar o perfil de expressão de 8156 janelas abertas de leitura (do inglês, ORFs) entre promastigotas, amastigotas derivados de lesão e amastigotas axênicos de *L. (L.) mexicana*. Os autores observaram que 288 genes (3,5%) apresentaram níveis de mRNAs com regulação diferencial entre promastigotas e amastigotas derivados de lesão. Curiosamente, a comparação dos resultados entre os promastigotas e os amastigotas axênicos resultou em somente 17 genes (0,2%) que satisfizeram os mesmos critérios estatísticos para regulação diferencial. A comparação entre os amastigotas derivados de lesão e os amastigotas axênicos resultou em 175 genes (2,1%) diferencialmente expressos. Todos estes dados sugerem que o reduzido número de genes regulados pode ser uma consequência do aumento na magnitude dos níveis de transcritos nas células sob condições axênicas (Holzer, McMaster et al., 2006).

Num outro trabalho, Leifso e colaboradores (2007) analisaram a expressão gênica diferencial entre promastigotas e amastigotas derivados de lesão de *L. (L.) major* através de microarranjos. Os níveis de expressão gênica demonstraram que 94% dos genes foram constitutivamente expressos em ambos os estágios de vida em *L. (L.) major*. Dentre os genes que foram expressos diferencialmente, com uma mudança na taxa de expressão superior a 2 vezes nos níveis de expressão, 115 genes (1,4%) foram preferencialmente expressos nos amastigotas e 125 genes (1,5%) foram preferencialmente expressos nos promastigotas (Leifso, Cohen-Freue et al., 2007).

Recentemente, Saxena e colaboradores (2007) estudaram o transcriptoma de *L.(L.) donovani* em escala genômica, baseados em microarranjos, e demonstraram pela primeira vez a progressão temporal das mudanças na abundância dos mRNAs durante a transição promastigota/amastigota num sistema *in vitro* que mimetizava este processo. As maiores descobertas neste estudo foram que a (i) expressão de um número substancial de genes são transientemente negativa ou positivamente regulados durante a diferenciação e (ii) que mudanças permanentes na expressão gênica ocorrem em tempos diferentes durante este fenômeno. Os dados demonstraram que aproximadamente 3% dos genes analisados foram superexpressos

nos amastigotas, 2,5% sofreram uma regulação negativa, e próximo de 3,5% dos genes sofreram mudanças transitórias durante o processo de diferenciação (Saxena, Lahav *et al.*, 2007).

De um modo geral, os vários trabalhos publicados utilizando diferentes abordagens para avaliação da expressão gênica no gênero *Leishmania* mostraram que os níveis dos transcritos ou dos produtos protéicos da maioria dos genes são constitutivamente expressos e que a taxa daqueles diferencialmente expressos oscila entre 3% e 9% (Walker, Vasquez *et al.*, 2006; Saxena, Lahav *et al.*, 2007; Guerfali, Laouini *et al.*, 2008; Li, Zhao *et al.*, 2008; Rochette, Raymond *et al.*, 2008). Deste modo, as análises através de ferramentas proteômicas, “microarrays” e “SAGE”, tanto das formas promastigotas quanto das formas amastigotas das diferentes espécies de *Leishmania*, demonstraram que o genoma parece ser constitutivamente expresso, com um número limitado de genes apresentando uma expressão diferencial durante o desenvolvimento. Porém, é sugestivo pensar que o controle específico dos processos biológicos relacionados pode estar vinculado aos aproximadamente 10% dos genes e proteínas diferencialmente expressos (o que representa cerca de 800 genes). O estudo destes genes de uma forma mais profunda pode levar à determinação de importantes genes regulatórios dos processos biológicos associados. Estes, por sua vez, podem ser usados como alvos para aplicação biotecnológica como, por exemplo, o desenvolvimento de novas drogas e vacinas.

Uma limitação de todas as abordagens das análises dos transcriptomas feitas até o momento é a quantidade de informações analisadas. Vários dos estudos de microarranjos e, principalmente de proteômica, usaram um número limitado de genes e proteínas. As principais desvantagens dessas técnicas são a dificuldade de análise de transcritos raros, a identificação de novos transcritos expressos ainda não catalogados, e a identificação de polimorfismos de nucleotídeos simples (do inglês, SNPs), que muitas vezes são de extrema importância para os processos celulares relacionados. Apesar do excelente desempenho na detecção de transcritos altamente abundantes, a “SAGE” (do inglês, Serial Analysis Gene Expression) envolve um sequenciamento relativamente limitado que não é tão robusto na detecção de transcritos raros (Wang, 2007). Isto é devido aos custos envolvidos com o extenso sequenciamento das bibliotecas SAGE pelo método de Sanger. Por causa destes

custos nenhuma biblioteca SAGE convencionalmente sequenciada exibe um número de “tags” saturantes, sugerindo uma incompleta representação do transcriptoma celular.

Em contraste, as tecnologias de sequenciamento de próxima geração (do inglês, NGS) oferecem uma redução substancial nos custos e um alto rendimento, com milhões de sequências obtidas numa única corrida (Torres, Metta *et al.*, 2008; Wilhelm, Marguerat *et al.*, 2008; Dong, Jensen *et al.*, 2009; Wang, Gerstein *et al.*, 2009; Wilhelm e Landry, 2009; Legendre, Audic *et al.*, 2010; Moe, Chung *et al.*, 2011). Um exemplo clássico desta poderosa ferramenta no contexto dos tripanosomas vem de um recente trabalho do grupo norte-americano da Universidade de Yale. Os autores do trabalho fizeram o mapeamento da expressão gênica “digital” e identificaram 1.114 novos transcritos, incluindo 103 RNAs não codificadores (do inglês, non-coding RNAs), determinaram precisamente os limites dos transcritos e a heterogeneidade dos sítios de “*trans* splicing” e poliadenilação, além de confirmar e corrigir muitas características das ORFs anotadas no genoma do *T. brucei* (Koley, Franklin *et al.*, 2010).

Poliadenilação estável da extremidade 3’ dos mRNAs em eucariotos

A poliadenilação de mRNAs eucarióticos é um evento biológico amplamente difundido e bastante conhecido, sendo a cauda poli(A) o alvo dos processos de purificação de mRNAs para uso em estudos de transcriptoma de eucariotos.

A transcrição pela RNA polimerase II produz diferentes classes de RNAs, incluindo os precursores dos RNAs mensageiros (pré-mRNAs) e os precursores de numerosas moléculas de RNAs estáveis, incluindo os pequenos RNAs nucleares (snRNAs) que atuam no processamento por “splicing” dos pré-mRNAs, os pequenos RNAs nucleolares (snoRNAs) com função na síntese de ribossomos, os miRNAs que regulam a tradução e a estabilidade de muitos mRNAs, além de muitos outros RNAs não codificadores (ncRNAs). Em eucariotos, a transcrição pela RNA Pol II é seguida co-transcricionalmente pela proteção da extremidade 5’ da mensagem contra a degradação por 5’ exonucleases. A adição de cauda poli(A) ao RNA é um processo comum em quase todos os organismos. A poliadenilação estável é uma modificação pós-transcricional que está associada com a extremidade 3’-madura da maioria dos mRNAs codificados no genoma nuclear dos eucariotos. A clivagem do pré-mRNA e

poliadenilação é um passo essencial para a formação da extremidade 3' de quase todos os transcritos codificantes de proteínas em eucariotos. A síntese funcional do mRNA envolve o processamento por clivagem endonucleolítica do mRNA nascente seguida pela adição de uma cauda poli(A) controlada por elementos *cis* ativadores no pré-mRNA circundando o sítio de clivagem.

Um complexo multi-protéico reconhece o sinal de poliadenilação, geralmente o hexanucleotídeo AAUAAA em metazoários, para dirigir o processamento do precursor (Proudfoot, 2011; Fuentes, Barrera *et al.*, 2012; Tian e Graber, 2012). Em outros eucariotos mais primitivos esta sequência consenso sofre pequenas variações, mas de um modo geral é conservada. Em *Trichomonas vaginalis* as análises de vários cDNAs sugeriram que um tetranucleotídeo UAAA poderia ser o sinal de poliadenilação (Fuentes, Barrera *et al.*, 2012). Nas células transfectadas com um mRNA exógeno contendo o motivo UAAA na 3'-UTR os transcritos deste gene heterólogo foram poliadenilados similarmente ao mRNA endógeno. Além do mais, a realocação da sequência UAAA redirecionou o processamento e a poliadenilação do mRNA repórter, confirmando o papel desta região como um sinal de poliadenilação canônico, similar à sequência consenso AAUAAA em metazoários (Fuentes, Barrera *et al.*, 2012).

Em Kinetoplastida, uma extensão de polipirimidinas de aproximadamente 300 nucleotídeos acima do sítio de “*trans* splicing” do gene adjacente no policistron marca o sinal de poliadenilação (Benz, Nilsson *et al.*, 2005). Análises de cDNAs de alguns genes em *T. brucei*, usando uma abordagem de RNA-Seq, mostraram que a poliadenilação é iniciada em um ou mais resíduos de adenosina (A) localizados de 80 a 140 nucleotídeos abaixo dessa extensão de polipirimidinas, similar aos sítios mais comuns previamente identificados (Benz, Nilsson *et al.*, 2005; Kolev, Franklin *et al.*, 2010). Em eucariotos em geral, estas caudas poli(A) estáveis participam de importantes fenômenos como, por exemplo, na exportação nuclear do mRNA maduro para o citosol, na iniciação da tradução e no auxílio da estabilidade da molécula (Wahle e Ruedgegger, 1999; Zhao, Hyman *et al.*, 1999; Edmonds, 2002; Proudfoot, 2004).

A compreensão dos eventos bioquímicos de poliadenilação da extremidade 3' madura dos mRNAs está em estágio relativamente avançado. Em mamíferos, cinco

proteínas separadas atuam em conjunto para mediar, primeiro, a clivagem da extremidade 3' do mRNA nascente e, segundo, a poliadenilação acoplada: (i) poli(A) polimerase (PAP), (ii) fator de especificidade de clivagem e poliadenilação (CPSF), (iii) fator de estimulação da clivagem (CstF), (iv e v) e fatores de clivagem I e II (CFIm e CFII_m). A composição das subunidades dos dois fatores multiméricos principais, CPSF e CstF, é bem conhecida, assim como são os seus contatos com a sequência consenso 3'-AAUAAA e os elementos abaixo do sítio de poliadenilação (DSE) ricos em U (Colgan e Manley, 1997; Zhao, Hyman *et al.*, 1999).

Os fatores homólogos em *Saccharomyces cerevisiae* estão parcialmente caracterizados e foram mostrados ser capazes de interagir com os elementos de sequência que constituem o sinal de poliadenilação em leveduras. Dois fatores multiméricos, chamados de fator de clivagem IA (CFIA) e fator de clivagem e poliadenilação (CPF), foram purificados e apresentaram subunidades compartilhadas com os fatores CstF e CPSF de mamíferos (Proudfoot e O'sullivan, 2002). Entretanto, ambos os fatores possuem subunidades adicionais que levaram à reavaliação das estruturas das subunidades das proteínas de mamíferos. CFIA tem 4 subunidades, duas das quais (Rna14p e Rna15p) possuem ortólogos em mamíferos, CstF77 e CstF64, respectivamente. De modo contrário, para as duas outras subunidades do fator CFIA, Pcf11p e Clp1p, não há ortólogos caracterizados no fator CstF, mas são encontrados (não tão bem caracterizados) no fator CFII_m de mamíferos (De Vries, Ruegsegger *et al.*, 2000). O fator CPF tem ao menos 15 polipeptídeos associados, incluindo todos os ortólogos das quatro subunidades do fator CPSF e a PAP (Pap1p em *S. cerevisiae*). Adicionalmente, a proteína Fip1p CPF-associada, originalmente identificada por sua interação com a Pap1p, parece ser a quinta subunidade, sem um ortólogo ainda identificado no componente CPSF de mamíferos. Ela pode atuar ligando a PAP ao CPSF no estágio inicial de poliadenilação. Assim, a PAP torna-se pronta para adicionar a cauda poli(A) à extremidade 3' recém produzida do mRNA (Kaufmann, Martin *et al.*, 2004). O Pti1p, outro fator associado ao CPF, apresentou significativa homologia com a subunidade Rna15p e pode também existir como uma variante do CFIA (Skaar e Greenleaf, 2002).

Parte da complexidade ainda não totalmente compreendida do fator CPF pode ser devida às outras classes de transcritos gerados pela Pol II (snRNAs e snoRNAs) fazerem uso do CFIA e CPF para gerar a extremidade 3' madura (Fatica, Morlando *et al.*, 2000; Morlando, Greco *et al.*, 2002). Se estes pequenos transcritos tem uma maquinaria dedicada à formação da extremidade 3' madura ou se eles essencialmente compartilham fatores envolvidos com esta atividade são questões ainda a serem respondidas. Contudo, as subunidades associadas ao CPF, Pti1p, Ref2p, Swd2p e Ssu72p estão envolvidas na formação da porção 3' madura de snRNAs e snoRNAs e podem existir como partes de um subcomplexo do fator CPF (Dheur, Vo Le *et al.*, 2003; Nedeá, He *et al.*, 2003; Steinmetz e Brow, 2003). Algumas subunidades adicionais do fator CPF estão provavelmente associadas com funções conectando a transcrição e o processamento da extremidade 3' dos mRNAs.

Uma barreira bioquímica nessas análises é que a natureza molecular da atividade das endoribonucleases que geram as extremidades 3' maduras dos mRNAs ainda são desconhecidas. Porém, outro estudo apontou a subunidade CPSF73 como uma provável candidata, uma vez que ela possui um potencial domínio nuclease que foi predito ser Zn²⁺-específico (Callebaut, Moshous *et al.*, 2002). Adicionalmente, a atividade de clivagem do complexo inteiro foi demonstrada requerer Zn²⁺ e, posteriormente, aminoácidos específicos neste domínio conservado, no ortólogo (Ysh1p) em *S. cerevisiae*, foram particularmente sensíveis à mutações (Ryan, Calvo *et al.*, 2004). A CFIm, uma proteína dimérica, foi demonstrada ser a reguladora de seleção do sítio de poliadenilação (Brown e Gilmartin, 2003). Embora ela seja um componente obrigatório do complexo de clivagem e poliadenilação, ela pode agir bloqueando seletivamente o uso de sítios inapropriados de poliadenilação. Com relação a isto, ela pode ser funcionalmente equivalente (mas não ortóloga) ao CFIB em *S. cerevisiae*, da mesma forma, sendo requerida para prevenir o uso de sítios inadequados de poliadenilação acima dos sítios canônicos de poliadenilação da extremidade 3'-madura dos mRNAs (Minvielle-Sebastia, Beyer *et al.*, 1998).

OBJETIVOS

Objetivo Geral

Produzir dados de transcriptoma em larga escala em duas espécies de *Leishmania* para estudar os fenômenos de poliadenilação de RNAs e sua relação com o controle da expressão gênica.

Objetivos Específicos

- Buscar nos dados de transcriptoma dos isolados ET e NSL de *L. (V.) braziliensis* e do isolado PH8 de *L. (L.) amazonensis* por *reads* poliadenilados na extremidade 3' correspondentes aos transcritos maduros das subunidades dos rRNAs;
- Identificar também em todas as bibliotecas transcriptômicas de *Leishmania* spp. a presença de *reads* poliadenilados nas extremidades 3' correspondentes os transcritos internos (truncados) das subunidades dos rRNAs e nas moléculas de mRNAs;
- Identificar nos dados de RNA-Seq a presença de caudas homo- e heteropoliméricas nos transcritos truncados de RNAs;
- Confirmar por RT-PCR os eventos de poliadenilação interna em rRNAs e mRNAs para alguns transcritos candidatos selecionados;
- Avaliar por RT-PCR a presença de transcritos truncados de rRNAs e mRNAs contendo caudas homo- ou heteropoliméricas no citoplasma de *L. (V.) braziliensis*;
- Buscar nas bibliotecas transcriptômicas de *Leishmania* spp. a presença das subunidades ortólogas às subunidades principais do exossoma de eucariotos superiores.

MATERIAIS E MÉTODOS

Camundongos

Foram utilizados camundongos fêmeas C57BL/6 (4 a 8 semanas de idade) criados e mantidos no Biotério Central da Universidade Federal de Ouro Preto (Biotério Central/NUPEB – UFOP, Ouro Preto/MG). Aos animais foram dados água e comida *ad libitum*. Todos os procedimentos envolvendo os animais estavam de acordo com o Comitê de Ética em Experimentação Animal da UFOP.

***Leishmania* spp., Isolados e Manutenção das Populações do Parasito**

Os parasitos da cepa PH8 (IFLA/BR/67/PH8) de *L. (L.) amazonensis* foram obtidos da coleção do banco de cepas congeladas em nitrogênio líquido do Laboratório de Imunoparasitologia coordenado pelo professor Dr. Luís Carlos Crocco Afonso. As cepas ET e NSL de *L. (V.) braziliensis* foram diretamente isoladas de biópsias de lesões da pele de pacientes humanos com casos confirmados de leishmaniose cutânea no Centro de Referência e Treinamento em Leishmanioses do Centro de Pesquisas René Rachou (CTRL-CPqRR/Fiocruz, Belo Horizonte/MG). Estas cepas foram mantidas em meio de cultura NNN/RPMI acrescido de 10% de soro fetal bovino e criopreservados em nitrogênio líquido pela Dra. Ana Lúcia Teles Rabelo. Estes isolados foram repassados ao Dr. Luís Carlos Crocco Afonso do Laboratório de Imunoparasitologia do Núcleo de Pesquisas em Ciências Biológicas da Universidade Federal de Ouro Preto-MG, aonde foram mantidos em meio Grace's com 10% de soro fetal bovino e caracterizados em suas respectivas atividades ectonucleotidásicas e infecciosidade em modelo murino. Os isolados foram, então, cedidos para os estudos de transcriptoma.

Para a manutenção *in vitro* das cepas de ambas as espécies de *Leishmania* os parasitos foram cultivados em Grace's Insect Medium (Sigma, Saint Louis, Missouri-EUA) suplementado com 10% de soro fetal bovino inativado (LGC Biotecnologia, Cotia, São Paulo-BRL), 2 mM de L-glutamina (Gibco BRL) e 100 U/mL de penicilina G potássica (USB Corporation, Cleveland, Ohio-EUA), pH 6.5 a 26 °C. As formas infectivas promastigotas metacíclicas foram obtidas por gradiente de centrifugação dos parasitos em fase logarítmica de crescimento (dia 5) num gradiente de enriquecimento com Ficoll 400 (Sigma, Saint Louis, Missouri-EUA). Brevemente, os parasitos foram lavados duas vezes em solução 0.9% NaCl, o pellet de células foi suspenso em Dulbecco's Minimal Essential Medium (Gibco/Life

Technologies, EUA), pH 7.2 e sobreposto na camada de Ficoll 10% em PBS. Esta preparação foi centrifugada a 1070 x g por 15 minutos a 25 °C, toda a fração do sobrenadante foi recolhida e novamente centrifugada a 1540 x g por 10 minutos a 4 °C. Os parasitos foram lavados duas vezes em solução 0.9% NaCl e posteriormente processados para extração de RNA (De Almeida Marques-Da-Silva, De Oliveira *et al.*, 2008).

Curva de Crescimento, Inóculo e Acompanhamento do Desenvolvimento de Lesão em Camundongos C57BL/6

A cultura foi iniciada com 1×10^5 *Leishmania*/mL. O crescimento foi acompanhado diariamente pela contagem dos parasitos até 10 dias após a preparação da cultura. Os dois isolados de *L. (L.) braziliensis* foram inoculados na pata traseira esquerda em grupos de camundongos C57BL/6 na concentração de 1×10^5 promastigotas metacíclicas/50 µL de PBS. O desenvolvimento da lesão foi acompanhado semanalmente com um *dial micrometer* (model 1015MA; L.S. Starret Co., Itu, SP, Brasil). Os resultados foram expressos como a diferença entre a média das medidas das patas infectadas e das contrapatas não infectadas (Afonso e Scott, 1993).

Fracionamento Celular e Purificação da Fração Polissomal

As frações citosólica e nuclear foram preparadas do isolado NSL de *L. (L.) braziliensis*. Inicialmente as células foram incubadas com cicloheximida (10 mg para 5×10^9 parasitos) e heparina (20 mg para 5×10^9 parasitos) por 10 minutos à temperatura ambiente. Os parasitos foram centrifugados a 7000 x g por 5 minutos a 4 °C, o sobrenadante foi descartado, as células foram lavadas duas vezes com tampão NKM (detalhes em Soluções e Tampões na seção Anexos) e novamente submetidas à centrifugação. A lise celular foi realizada pela suspensão do pellet em 15,75 mL de tampão de lise hipotônico brando (detalhes em Soluções e Tampões na seção Anexos). As células foram permeabilizadas pela adição de 1,75 mL de detergente Nonidet P-40 à concentração final de 1% e a ruptura foi monitorada por visualização em microscópio ótico. A lise celular foi interrompida pela adição de sacarose à concentração final de 0,25 M e centrifugação do lisado a 10.000 x g por 10 minutos a 4 °C. O pellet foi rotulado como fração nuclear, o sobrenadante foi transferido para um novo tubo e referido como fração citosólica. Esta fração citosólica foi novamente

centrifugada a 10.000 x g por 30 minutos a 4 °C para sedimentação das mitocôndrias. O sobrenadante foi coletado e referido como fração pós-mitocondrial (correspondente ao RNA citosólico solúvel total sem a fração nuclear).

Para o enriquecimento do extrato polissomal a fração pós-mitocondrial foi submetida ao processo de ultra-centrifugação (Beckman Coulter Corporation). Foi acrescentado ao fundo do tubo de centrífuga 2 mL de sacarose 2 M (devidamente acrescida de cicloheximida, heparina e β -mercaptoetanol; detalhes em Soluções e Tampões na seção Anexos) e, cuidadosamente adicionado por cima, 9 mL da fração pós-mitocondrial. A amostra foi centrifugada (rotor SW41 Ti) a 271.000 x g por 2 horas a 4 °C. Imediatamente após o término da centrifugação, a fração pós-polissomal (superior) referente à fração citosólica solúvel foi coletada em um novo tubo. A sacarose foi desprezada e o pellet contendo a fração polissomal enriquecida foi suspenso em 750 μ L de Trizol (Ambion/Life Technologies, EUA). Todas as frações foram armazenadas em freezer a -80 °C. As frações nuclear, polissomal e pós-polissomal foram posteriormente processadas pelo método do Trizol para extração das moléculas de RNA. Todos estes procedimentos foram realizados como previamente descrito (Alves, Avila *et al.*, 2010).

Os extratos nuclear e citosólico foram verificados quanto à eficiência do fracionamento celular pela dosagem da atividade da enzima citosólica Glicose-6-Fosfato-Desidrogenase (G6PDH Assay Kit, BioVison Life Science, San Francisco/CA, EUA) e pela quantificação de DNA (Quant-iT™ PicoGreen dsDNA Reagent and Kits, Invitrogen/Life Technologies, Carlsbad/CA, EUA).

Ensaio de Atividade Enzimática da Glicose-6-Fosfato-Desidrogenase

A glicose-6-fosfato-desidrogenase (G6PDH) é uma enzima citosólica da via pentose fosfato, uma rota metabólica que fornece energia reducionante às células pela manutenção dos níveis da coenzima nicotinamida adenina dinucleotídeo fosfato (NADPH). O ensaio de dosagem da atividade enzimática da G6PDH foi realizado para avaliar o grau de pureza do processo de fracionamento celular. Foi utilizado o kit Glucose-6-Phosphate Dehydrogenase Assay (BioVison, San Francisco/CA, EUA) por ser um método simples, sensível e rápido. No ensaio, a glicose-6-fosfato é oxidada com a geração de um produto no qual é utilizado para converter uma sonda incolor num produto intensamente colorido que absorve luz em 450 nm.

Os extratos obtidos a partir do fracionamento celular foram submetidos ao teste para avaliação da presença da enzima G6PDH. Os testes foram realizados em placas de 96 poços, em duplicata, em três volumes diferentes para cada amostra – 5 μL , 15 μL e 25 μL – e o volume final foi ajustado para 50 μL com o tampão de atividade. Esta variação nas dosagens foi justamente para garantir que as leituras ficassem dentro da faixa linear. O controle positivo foi diluído 100 vezes em tampão de atividade e, da mesma forma que para as amostras, três volumes diferentes foram testados – 1 μL , 5 μL e 10 μL – e o volume final foi ajustado para 50 μL com o tampão de atividade. Para a curva padrão foram adicionados 2, 4, 6, 8 e 10 μL da solução padrão de NADH 1.25 mM, em duplicata, em placa de 96 poços para gerar 2,5; 5,0; 7,5; 10,0 e 12,5 nmol/poço. O volume final foi ajustado para 50 μL com tampão de atividade. Então foi adicionado 50 μL do mix de reação aos poços das amostras, do controle positivo e da curva padrão. O mix de reação continha: 46 μL de tampão de atividade da G6PDH, 2 μL de substrato e 2 μL de reagente colorimétrico. As placas foram incubadas a 37 °C (protegidas da luz) e as medidas de absorvância (450 nm) foram capturadas nos tempos 0, 2, 5, 10, 15, 20, 25 e 30 minutos do desenvolvimento da reação.

Extração de RNA Total

O Trizol (Ambion/Life Technologies, EUA) é um reagente monofásico de fenol, isotiocianato de guanidina e outros componentes, designado para isolar com alta qualidade a diversidade de RNAs (moléculas grandes e pequenas) de amostras líquidas de uma infinidade de tipos celulares de modo rápido, fácil e eficiente. Para este propósito, as células de *Leishmania* em suspensão na cultura foram coletadas por centrifugação (1540 x g; 10 minutos; 4 °C) e o sobrenadante descartado. O reagente de Trizol foi adicionado ao pellet de células na proporção de 3:1 (750 μL de Trizol para cada 250 μL de suspensão celular na concentração de 4×10^8 *Leishmania*/mL) e pipetado para cima e para baixo várias vezes para promover a lise celular. O homogenato foi incubado por 5 minutos em temperatura ambiente para completa dissociação dos complexos nucleoprotéicos. Foram adicionados 200 μL de clorofórmio (para cada 750 μL de Trizol), agitado vigorosamente em vortex por 15 segundos, incubado por 10 minutos em temperatura ambiente e centrifugado para separação das fases orgânicas (12000 x g; 15 minutos; 4 °C). Como o RNA permanece exclusivamente na fase superior aquosa esta foi coletada para os passos

subsequentes. Foram adicionados à fase aquosa 500 µL de isopropanol 100% (gelado; para cada 750 µL de Trizol), incubada por 10 minutos em temperatura ambiente e novamente centrifugada (12000 x g; 10 minutos; 4 °C). O sobrenadante foi descartado e o pellet de RNA foi lavado com 1 mL de etanol 75% (gelado; para cada 750 µL de Trizol). A amostra foi centrifugada a 7500 x g por 5 minutos a 4 °C e o sobrenadante descartado. O pellet de RNA foi seco em temperatura ambiente por 10 minutos, suspenso em 20 µL de H₂O DEPC e incubado por 10 minutos a 60 °C. As amostras foram estocadas no freezer -80 °C e posteriormente quantificadas, avaliadas por eletroforese microcapilar e utilizadas para a síntese de cDNAs. O mesmo procedimento foi realizado para as frações celulares.

Extração de RNAs poli(A)⁺

Os RNAs poli(A)⁺ foram extraídos para a construção das bibliotecas transcriptômicas e para os ensaios de RT-PCR. Para esta finalidade utilizamos o FastTrack MAG mRNA Isolation kit (Invitrogen/Life Technologies, Carlsbad/CA, EUA). O isolamento e captura das moléculas de RNAs poli(A)⁺ foi realizada através dos “beads” magnéticos conjugados com oligo(dT). Todos os procedimentos foram realizados de acordo com as instruções do manual do fabricante, com exceção de algumas pequenas modificações realizadas a fim de aumentar o rendimento final. Todas as extrações foram realizadas na proporção de 1 x 10⁸ *Leishmania* para 100 µL de “beads” magnéticos. Esta proporção foi a que forneceu os melhores resultados, otimizando o processo de extração.

Brevemente, as células dos parasitos foram coletadas e incubadas com o tampão de lise celular (Lysis buffer L4 e inibidor de RNase não protéico – ribonucleoside vanadyl complex 10 mM). O lisado foi passado em uma seringa de 1 mL (agulha calibre 21 G) no mínimo 20 vezes para garantir a completa fragmentação do DNA. Uma alíquota de 100 µL de “beads” magnéticos foi separada e estes foram lavados duas vezes em Wash buffer W7. Posteriormente, aos “beads” magnéticos foi adicionado o extrato celular juntamente com o Binding buffer B6 e incubados a 70 °C por 5 minutos. A mistura (chamada a partir daqui de amostra) foi transferida para um agitador e incubada por 10 minutos em temperatura ambiente. A amostra foi colocada no separador magnético de partículas (SMP) e incubada por 2 minutos (ou até a completa separação magnética dos “beads” – inspeção visual até os “beads”

aderirem à parede do tubo). O sobrenadante foi coletado e novamente incubado com uma nova alíquota de “beads” magnéticos para captura dos RNAs “raros” (pequenos RNAs poli(A)⁺; o processo foi repetido para estes RNAs raros). Os “beads” magnéticos que ficaram no tubo foram lavados com 500 µL de Wash buffer W6 contendo 1 µL de DNase I Amplification Grade (Invitrogen/Life Technologies, Carlsbad/CA, EUA), misturados levemente e incubados a 25 °C por 10 minutos. A amostra foi novamente inserida no SMP, o sobrenadante desprezado e os “beads” lavados três vezes em Wash buffer W7. Após a última lavagem a amostra foi retirada do SMP e os “beads” foram homogeneizados com 20 µL de H₂O DEPC para eluição dos RNAs poli(A)⁺. A amostra foi incubada a 37 °C por 5 minutos e novamente inseridas no SMP. O sobrenadante com os RNAs poli(A)⁺ foi coletado e armazenado no freezer a -80 °C.

Quantificação de RNA e DNA por Método Fluorescente

Quantificamos os ácidos nucleicos através de um reagente fluorescente. Para a dosagem de DNA foi utilizado Quant-iT™ PicoGreen e para a dosagem de RNA utilizamos o Quant-iT™ RiboGreen (Invitrogen/Life Technologies, Carlsbad/CA, EUA). O processo de quantificação é idêntico para as duas moléculas. Para quantificação seguimos os passos recomendados no manual do fabricante.

Sucintamente, a etapa inicial na quantificação é o preparo da solução de trabalho diluindo 20 vezes o TE buffer concentrado (200 mM Tris-HCl, 20 mM EDTA, pH 7,5). Esta solução é usada para a diluição do RiboGreen/PicoGreen fluorescente e do RNA/DNA. A quantificação foi ajustada para um volume de 200 µL em microplaca de 96 poços e excitação de fluorescência a 485 nm, sendo a emissão em 525 nm. A curva padrão foi montada (em duplicata) por diluição seriada da solução padrão com os pontos da curva compreendidos entre 50 e 0,78125 ng. As amostras foram diluídas 100 vezes (1 µL da amostra + 99 µL do TE buffer), aplicadas na placa e em seguida foi adicionado 100 µL do reagente fluorescente (preparado em TE buffer) em cada poço (das amostras e da curva padrão). A placa foi incubada por 5 minutos em temperatura ambiente e protegida da luz e a leitura realizada em leitora de microplacas (Synergy HT Multi-Mode Microplate Reader – BioTek, Winooski/VT, EUA).

Eletroforese Microcapilar das Amostras de RNA Total e RNA poli(A)⁺

Para avaliar a integridade das amostras de RNA total ou das amostras de RNA poli(A)⁺, incluindo a presença ou ausência de produtos de degradação, foi utilizado o kit Agilent RNA 6000 Pico (Agilent Technologies, Waldbronn, Alemanha). O ensaio eletroforético é baseado nos princípios de eletroforese em gel, porém, transferidos para o formato de “chip” (“on-chip gel electrophoresis”). Isto reduz dramaticamente o tempo de separação bem como o consumo de amostras e de reagentes. O sistema fornece uma informação automatizada de quantificação e de tamanho em formato digital. As biomoléculas carregadas são separadas por tamanho no qual o corante é intercalado entre as fitas de DNA ou RNA e estes complexos são detectados por fluorescência induzida a laser. Os dados são traduzidos em imagens “gel-like” (bandas) e eletroferogramas (picos). Os tempos de migração são medidos para cada amostra e o tamanho é calculado. O marcador é usado como uma referência interna para alinhar o padrão com as amostras. Uma grande vantagem da técnica é que para os ensaios com RNA, a taxa de ribossomal é determinada após a eletroforese.

As etapas do processo serão sucintamente resumidas a seguir. Para o preparo do gel 550 µL do RNA 6000 Pico Gel Matrix foram pipetados num tubo Spin Filter e centrifugados a 1500 x g por 10 minutos em temperatura ambiente. O gel filtrado foi estocado a 8 °C em alíquotas de 65 µL. Para a eletroforese microcapilar das amostras a primeira etapa é o preparo do Gel-Dye mix. É necessário deixar todos os reagentes equilibrarem em temperatura ambiente por 30 minutos. O RNA 6000 Pico Dye foi homogeneizado e 1 µL foi adicionado a uma alíquota de 65 µL do gel filtrado. O mix foi homogeneizado e centrifugado a 13000 x g por 10 minutos em temperatura ambiente. Um “chip” RNA 6000 Pico foi colocado na Priming Station e 9 µL do Gel-Dye mix foram pipetados no poço marcado com o símbolo G. A tampa da Priming Station foi fechada, o êmbolo da seringa foi ajustado para a posição de 1 mL e uma pressão (para espalhar o mix no “chip”) foi exercida por 30 segundos e então a trava foi liberada, desfazendo a pressão. Mais 9 µL foram aplicados em cada um dos outros dois poços marcados com G. No poço marcado com CS foi aplicado 9 µL do RNA 6000 Pico Conditioning Solution e 5 µL do RNA 6000 Pico Marker foram aplicados nos 11 poços de amostra e no poço do Ladder. Em seguida 1 µL do RNA 6000 Pico Ladder foi adicionado no respectivo poço. As amostras foram previamente

desnaturadas a 70 °C por 2 minutos e 1 µL de cada amostra foi aplicada. Nos poços não utilizados foi adicionado 1 µL de H₂O RNase-free. O “chip” foi fixado no IKA Vortexer e agitado por 1 minuto a 2400 rpm. O “chip” foi transferido para o Bioanalyzer e a corrida foi iniciada utilizando uma das duas configurações possíveis – Eukaryotic RNA Total Pico Series II ou Eukaryotic mRNA Pico Series II – dependendo do tipo da amostra.

Reação de Transcriptase Reversa [oligo(dT) RT-PCR] e Reação em Cadeia da Polimerase (PCR)

Para as reações de síntese da primeira fita de DNA complementar (cDNA) utilizamos a SuperScript™ III Reverse Transcriptase (Invitrogen/Life Technologies, Carlsbad/CA, EUA). O volume final da reação foi de 20 µL e a massa de RNA total usada como “template” foi sempre de 500 ng. Usamos um “primer” não ancorado para a transcrição reversa sintetizado com uma sequência adaptadora de 22 nucleotídeos na extremidade 5’ e uma cauda oligo(dT)₁₀ na ponta 3’ (ver detalhe do “primer” utilizado no subitem Lista de Oligonucleotídeos Utilizados Neste Trabalho). Também definimos uma concentração fixa do “primer” (500 ng) para todas as reações. À reação foi adicionado 1 µL de dNTP mix 10 mM, e então previamente aquecida a 65 °C por 5 minutos e incubada no gelo por 1 minuto para anelamento do “primer”. Em seguida foi adicionado 4 µL do tampão 5X First-Strand, 1 µL de DTT 100 µM, 1 µL RNaseOUT Recombinant RNase Inhibitor (Invitrogen/Life Technologies, Carlsbad/CA, EUA) e 1 µL da enzima SuperScript™ III RT. A reação foi incubada a 50 °C por 60 minutos para amplificação no Termociclador Dx 96-Well (Applied Biosystems/Life Technologies, Foster City/CA, EUA) e inativada pelo aquecimento a 70 °C por 15 minutos. Para remover o RNA complementar ao cDNA foi adicionado 1 µL (2 unidades) de RNase H (Invitrogen/Life Technologies, Carlsbad/CA, EUA) e incubado a 37 °C por 20 minutos.

As PCRs foram realizadas com o kit GoTaq Green Master Mix (Promega Corporation, Madison/WI, EUA). Todas as reações foram realizadas com uma quantidade de cDNA “template” padronizada (20 ng). A concentração final dos “primers” utilizados foi de 500 nM (ver detalhe do “primer” utilizado no subitem Lista de Oligonucleotídeos Utilizados Neste Trabalho). As etapas da reação foram:

95 °C por 10 minutos para a desnaturação inicial; 30 ciclos de amplificação a 95 °C por 1 minuto para desnaturação; 55 °C (“primer” específico para mRNA) e 60 °C (“primer” específico para rRNA) por 30 segundos para anelamento dos “primers”; amplificação a 72 °C por 2 minutos; etapa de amplificação final a 72 °C por 10 minutos.

Purificação de DNA (amplicons RT-PCR) do Gel de Agarose

Após a PCR, o volume total da reação (25 µL) foi aplicado num gel de agarose 1,2% e separado por eletroforese a 120 Volts por 60 minutos em tampão TAE (detalhes em Soluções e Tampões na seção Anexos). A purificação de todos os produtos de amplificação (banda forte e “smear”) foi realizada usando o GenElute Gel Extraction Kit (Sigma, Saint Louis/Missouri, EUA). O kit permite a rápida purificação de fragmentos de DNA de 50 pb a 10 kb pela excisão das bandas de DNA do gel de agarose e consequente solubilização do gel. A solução de solubilização também contém um indicador de pH que permite que o fragmento do gel seja visualizado com facilidade e indica se o pH está ideal para a ligação do DNA à coluna. Os fragmentos de DNA extraídos são seletivamente adsorvidos na membrana de sílica e os contaminantes são removidos por centrifugação. Finalmente, o DNA ligado à membrana é coletado com a solução de eluição por centrifugação.

As etapas foram seguidas de acordo com o manual de instruções do fabricante e serão comentadas resumidamente. Todas as centrifugações foram realizadas em 12000 x g. Os fragmentos de interesse foram excisados do gel de agarose, pesados e adicionados 3 volumes de solução de solubilização do gel. Em outras palavras, para cada 100 mg de gel foram adicionados 300 µL de solução de solubilização. A mistura foi incubada em banho-maria a 60 °C por 10 minutos e em seguida homogeneizada para completa dissolução do gel. As colunas contendo uma membrana de sílica foram preparadas previamente pela lavagem com solução de preparação da coluna e centrifugadas por 1 minuto. À solução do gel dissolvido foi adicionado 1 volume de isopropanol 100% e a mistura aplicada na coluna de ligação e centrifugada por 1 minuto. A coluna foi lavada duas vezes com 700 µL de solução de lavagem e centrifugada por 1 minuto. A coluna de ligação foi transferida para um novo tubo de coleta. Finalmente, foi adicionado 20 µL de solução de eluição

previamente aquecido a 65 °C. O DNA foi eluído por centrifugação por 1 minuto e armazenado em freezer -20 °C.

Clonagem dos Produtos de PCR Purificados

Após a purificação dos produtos de PCR do gel de agarose prosseguimos com a clonagem dos fragmentos de DNA. Utilizamos o CloneJET™ PCR Cloning Kit (Fermentas, Burlington/ON, Canadá). Como as PCRs foram realizadas com a GoTaq DNA polimerase os produtos gerados possuem uma 3'-dA “overhang” pelo fato de a enzima não possuir uma atividade revisora 3'→5' que remove as extremidades despareadas. Deste modo, optamos pelo método de clonagem “Sticky-End” descrito no manual do fabricante. O vetor pJET1.2/blunt recircularizado expressa um gene letal (eco47IR) após a transformação e não pode ser propagado. Este gene letal é rompido pela ligação de um inserto de DNA no sítio de clonagem. O vetor também contém o gene β-lactamase, que confere resistência à ampicilina, selecionando e mantendo as células de *Escherichia coli* recombinantes. Como resultado, somente os clones recombinantes contendo o inserto aparecem nas placas de cultura. As etapas brevemente descritas abaixo foram seguidas de acordo com as informações contidas no manual do fabricante.

A enzima DNA Blunting remove a 3'-overhang e preenche a extremidade complementar 5'-overhang. A reação de “Blunting” foi preparada em banho de gelo. Num tubo de PCR foi adicionado 10 µL do tampão de reação 2X, 5 µL do produto de PCR purificado do gel, 1 µL da enzima DNA Blunting e 2 µL de H₂O nuclease-free. A mistura foi homogeneizada suavemente e centrifugada por 5 segundos para sedimentar as gotículas de reação. A reação de “Blunting” foi incubada a 70 °C por 5 minutos e colocada novamente no gelo. Ainda no gelo, foi adicionado à mistura da reação 1 µL do vetor pJET1.2/blunt (50 ng/µL) e 1 µL da enzima T4 DNA Ligase (volume final = 20 µL). A mistura de ligação foi incubada por 30 minutos em temperatura ambiente e em seguida usada diretamente para a transformação de células de *E. coli* DH5 alfa competentes.

Preparo de Células Competentes por Cloreto de Cálcio

Uma colônia de *E. coli* DH5 alfa foi inoculada em 4 mL de meio Luria-Bertani (detalhes do meio LB em Soluções e Tampões na seção Anexos) e crescida

“overnight” a 37 °C sob constante agitação. No dia seguinte 2 mL da cultura de bactérias foram inoculados em 100 mL de meio LB (sem antibiótico) e incubados a 37 °C sob agitação até ter atingido uma densidade ótica de 0,5. A cultura foi centrifugada a 4000 x g por 10 minutos a 4 °C. O “pellet” de células foi colocado em banho de gelo, suspenso em 40 mL de CaCl₂ 0,1 M gelado e as células novamente incubadas por 1 hora no gelo. Após a centrifugação a 4000 x g por 10 minutos a 4 °C as células foram suspensas em 1 mL de CaCl₂ 0,1M gelado. Por fim, as células foram estocadas em freezer -80 °C em alíquotas de 100 µL com 20% de glicerol estéril.

Transformação das Células Competentes por Choque Térmico

Para cada clonagem foi descongelado um tubo de células de *E. coli* DH5 alfa competentes. Os 20 µL da reação de ligação foram adicionados aos 100 µL de células competentes, a mistura foi homogeneizada e incubada em gelo por 45 minutos. Em seguida as células sofreram um choque térmico a 42 °C por 90 segundos e imediatamente foi adicionado 1 mL de meio LB. As células foram incubadas a 37 °C por 40 minutos e plaqueadas em meio LB-ágar (ampicilina 50 µg/mL). As placas foram incubadas “overnight” em estufa a 37 °C. Após visualização das colônias transformantes, 24 clones de cada clonagem foram coletados com o auxílio de palitos de dentes autoclavados e inoculados em 0,1 mL de meio LB contendo ampicilina (50 µg/mL) em placas de 96 poços para crescimento e posterior estocagem com adição de glicerol (20% v/v) e congelamento a -80 °C.

Miniprep de DNA Plasmidial em Placa de 96 Poços

1) Preparação da cultura líquida (placa de 96 poços)

Para cada placa com *E. coli* DH5 alfa transformada selecionamos 24 clones. Foram utilizadas microplacas de 96 poços para inóculo e extração de DNA. As colônias de DH5 alfa, contendo o plasmídeo com o inserto a ser sequenciado, foram repicadas e armazenadas em placas com meio líquido LB e 20% de glicerol em freezer -80 °C. Uma réplica de cada placa, a partir dos estoques em glicerol, foi produzida em uma nova placa de 96 poços com 2 mL de meio líquido LB-ampicilina (50 µg/mL) e crescida “overnight” a 37 °C em constante agitação.

2) Extração de DNA plasmidial (placa de 96 poços)

As placas foram centrifugadas a 4000 x g por 6 minutos a 4 °C, o sobrenadante foi desprezado e a placa foi invertida sobre papel absorvente para que o “pellet” secasse durante 5 minutos. O “pellet” de cada poço na placa foi homogeneizado com 240 µL de tampão GTE (detalhes em Soluções e Tampões na seção Anexos). O processo de centrifugação e secagem do “pellet” foi repetido. Foi adicionado 85 µL da mistura (para 2 placas: 17,6 mL de tampão GTE e 1,1 mL de RNase – 10mg/mL) em cada poço, homogeneizado por pipetagem para suspensão do “pellet” e 60 µL desta solução foi transferida para uma nova placa com fundo em U. Para cada poço foi adicionado 60 µL de solução de lise NaOH/SDS (detalhes em Soluções e Tampões na seção Anexos). A microplaca foi selada com um filme adesivo (Axygen), misturada por inversão 10 vezes, incubada por 10 minutos em temperatura ambiente e centrifugada rapidamente. Foi adicionado 60 µL de acetato de potássio 3 M (pH 4,8), mantido a 4 °C e repetido o processo de incubação. O adesivo foi removido e a microplaca foi colocada numa estufa a 90 °C por 30 minutos. A placa foi resfriada no gelo por 10 minutos e centrifugada a 4000 x g por 4 minutos a 20 °C.

Para filtrar e clarear as amostras de DNA utilizamos uma placa de 96 poços com filtro MILLIPORE, de forma alinhada, acoplada em cima de uma nova placa de 96 poços com fundo em “U”. Todo o volume do sobrenadante foi transferido para o filtro e o conjunto foi centrifugado a 4000 x g por 4 minutos a 20 °C. O filtro foi removido e adicionado 110 µL de isopropanol absoluto por poço ao filtrado. A placa foi selada e homogeneizada por inversão e posteriormente centrifugada a 4000 x g por 5 minutos a 20 °C. O sobrenadante foi removido e adicionados 200 µL de etanol 70% gelado. A placa foi centrifugada a 4000 x g por 5 minutos a 20 °C. O sobrenadante foi novamente descartado e a microplaca foi centrifugada na posição invertida a 250 x g por 3 minutos a 20 °C para completa remoção dos líquidos. Por fim, a placa foi seca por 1 hora em temperatura ambiente, o DNA precipitado foi suspenso em 10 µL de H₂O nuclease-free e a placa foi estocada em freezer -20 °C. A qualidade do DNA extraído em microplacas foi verificada por eletroforese em gel de agarose 1,2% e separado por eletroforese a 120 Volts por 60 minutos em tampão TAE, visualizado por coloração por brometo de etídio sob luz ultravioleta.

Reação de Sequenciamento, Precipitação e Sequenciamento Automático do DNA

Para determinação das sequências de DNAs purificados foi utilizado um protocolo adaptado a partir do original descrito previamente (Sanger, Nicklen *et al.*, 1977). Todos os procedimentos foram realizados utilizando placas de 96 poços. As reações de sequenciamento automático foram preparadas em placas de fundo cônico utilizando o reagente BigDye Terminator Cycle Sequencing Ready Reaction v3.1 (Applied Biosystems/Life Technologies, Foster City/CA, EUA). As reações continham no volume final (20 µL) 2 µL de BigDye, 1 µL de iniciador (pJET 1.2 forward ou pJET 1.2 reverse sequencing primer) na concentração de 10 µM (10 pmoles), 300 – 500 ng de DNA plasmidial e 6 µL de tampão do Big Dye. As reações foram incubadas por 1 minuto a 95 °C e submetidas em termociclador PCR System 9700 (PerkinElmer, Waltham/MA, EUA) a 40 ciclos com desnaturação a 95 °C por 30 segundos, anelamento a 50 °C por 30 segundos, e extensão a 60 °C por 4 minutos.

Após a marcação foram adicionados, em sequência a cada amostra, 80 µL de isopropanol 75%. A placa foi selada com um filme adesivo (Axygen) e a mistura foi feita por inversão seguida de incubação por 15 minutos à temperatura ambiente e centrifugação a 2500 x g por 45 minutos. O sobrenadante foi descartado e seu resquício retirado por inversão da placa de 96 poços sobre papel absorvente. Foram adicionados 200 µL de etanol 70% e procedeu-se nova centrifugação a 2500 x g por 6 minutos. O sobrenadante foi descartado e a lavagem com etanol 70% foi repetida. O DNA precipitado foi então solubilizado em tampão formamida Hi-Di (Applied Biosystems/Life Technologies, Foster City/CA, EUA) e processado no sequenciador automático ABI PRISM 3100 Genetic Analyzer (Applied Biosystems/Life Technologies, Foster City/CA, EUA). Estes procedimentos foram realizados no Laboratório de Parasitologia Molecular da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRP/USP) sob coordenação da professora Dra. Angela Kaysel Cruz.

Tabela 1. Lista de Oligonucleotídeos Utilizados Neste Trabalho.

Nome do oligo	Gene	Posição	Direção	Aplicação	Sequência 5' → 3'
RV_RT-PCR_adapter_poliT	-	-	Reverso	dT-RT-PCR	GACTCGAGTCGACATCGATGAT ₁₀
RV_general adapter	-	-	Reverso	dT-RT-PCR	GACTCGAGTCGACATCGATGAT
FW_Lb_PRO_NSL_18S	SSU 18S rRNA	1260	Direto	dT-RT-PCR	AGTCCACACTGCAAACGATG
FW_Lb_PRO_NSL_28S beta	LSU 28S rRNA	858	Direto	dT-RT-PCR	AGGGTGGTGTGCTTTTTCTG
FW_Lb_PRO_NSL_eiF5a2	Eukaryotic initiation factor 5a putative	72	Direto	dT-RT-PCR	TGCTGGTGCCTTGAAGAAG
FW_Lb_PRO_NSL_60S ribosomal protein	60S ribosomal protein L5 putative	115	Direto	dT-RT-PCR	CAGGACAAGACGAAGTTTGG
pJET1.2 Forward Sequencing Primer	-	310	Direto	Clonagem-PCR	CGACTCACTATAGGGAGAGCGGC
pJET1.2 Reverse Sequencing Primer	-	428	Reverso	Clonagem-PCR	AAGAACATCGATTTTCCATGGCAG

Construção e Sequenciamento das Bibliotecas Transcriptômicas de RNA-Seq

Para o sequenciamento de todas as bibliotecas transcriptômicas utilizamos uma tecnologia de sequenciamento maciçamente paralelo (do inglês, MPS) de próxima geração (do inglês, NGS). A plataforma escolhida foi a de pirosequenciamento desenvolvida pela companhia Roche – 454 GS FLX Titanium Series. Todos os procedimentos experimentais – construção das bibliotecas de cDNAs, PCR em emulsão e sequenciamento – foram realizados de acordo com as instruções do fabricante. Estes procedimentos foram realizados no Centro Avançado em Tecnologia Genômica (CATG) do Instituto de Química da Universidade de São Paulo (IQ/USP) sob coordenação do professor Dr. Sérgio Verjovski-Almeida. A execução destas etapas foi auxiliada pelos doutorandos, do Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo (IQ/USP), Murilo Sena Amaral e Felipe Cesar Ferrarezi Beckedorff.

Análises de Bioinformática

1) Mapeamento de rRNAs

Todos os dados obtidos a partir do sequenciamento das bibliotecas transcriptômicas (RNA-Seq) foram analisados na plataforma do software CLC Genomics Workbench (Versão 5.0.1). Para gerar os dados das análises genômicas usamos como “template” as sequências depositadas no banco de dados de sequências genômicas dos tripanosomatídeos (www.tritrypdb.org). Para o mapeamento dos *reads* que alinharam no “cluster” dos genes de rRNAs utilizamos como “template” o cromossomo 27 da *L. (L.) major* (cepa referência – MHOM/IL/80/Friedlin, zymodeme MON-103). A escolha das sequências de *L. (L.) major* como referência foi baseada no fato de que até o presente momento o genoma de *L. (L.) amazonensis* ainda não está disponível e mesmo com mais de 8000 genes descritos no genoma de *L. (V.) braziliensis*, seu genoma ainda não está completamente finalizado. Isto é evidenciado pelo alto número de “gaps” e “contigs” em relação ao número predito de cromossomos (Peacock, Seeger *et al.*, 2007). Adicionalmente, o “cluster” de genes ribossomais (rDNA) para *L. (V.) braziliensis* não está predito na última liberação do TritrypDB (www.tritrypdb.org; release 3.3 – August 19, 2011).

O primeiro passo foi mapear todos os *reads* no “template” de referência. Isto é feito para criar um alinhamento de todos os *reads*. Por último, fizemos a montagem *de novo* dos *reads*. Resumidamente, isto significa que tínhamos a intenção de que o CLC Genomics Workbench criasse sequências consensos mais contínuas (longas) a partir dos *reads* relativamente curtos. O processo de montagem agrupou todos os *reads* em sequências chamadas de “contigs”. O resultado da montagem é uma tabela de mapeamento onde cada linha na tabela representa um “contig”. O processo final na análise foi o mapeamento dos *reads* (“contigs”) ao genoma de referência (cromossomo 27 de *L. (L.) major*). Ao final do mapeamento foram gerados os seguintes resultados:

- Lista dos *reads* não mapeados: apresentam os *reads* que não mapearam no “template” (genoma de referência) por não ter atendido os critérios de mapeamento;
- Relatório: contém as informações sobre o mapeamento. Mais importante, ele mostra o número de *reads* que alinharam com a sequência de referência;
- Mapeamento: O mapeamento em si mostra o alinhamento de todos os *reads* para com a referência genômica e inclui os “quality scores” de cada base.

Assim, para o mapeamento dos *reads* no cluster de rRNA todos os *reads* foram filtrados de acordo com parâmetros menos estridentes – 50% de cobertura e 80% de similaridade. Nestas condições, o CLC Genomics Workbench realizou um alinhamento local dos *reads* com a sequência de referência. Com o parâmetro de cobertura ajustado para 0,5 (50%) aumentamos as possibilidades de que mais *reads* com as extremidades 3’ desalinhadas fossem filtrados no mapeamento por permitirmos que no mínimo 50% de cada *read* alinhasse com o “template”. À outra metade foi permitido divergir da referência. Estes critérios adotados possibilitaram uma condição mais “relaxada” que foi importante por permitir com que os *reads* contendo as extensões de poli(A)⁺ na extremidade 3’ fossem identificados durante o mapeamento. Da mesma forma, o número de “mismatches” dos *reads* foi levado em consideração. Estes “mismatches” são internos ao alinhamento local, sendo as extremidades desalinhadas desconsideradas para este critério. Com o parâmetro de similaridade de 0,8 (80%) aceitamos que a região do *read* alinhada ($\geq 50\%$) deveria parear com 80% de similaridade à sequência do “template”.

2) Mapeamento das subunidades do exossoma de degradação de RNAs

O exossoma, maquinaria ligada à degradação de RNAs, é altamente conservado entre os eucariotos e é formado por nove subunidades principais que são compartilhadas pelas formas citoplasmática e nuclear do complexo (Chen, Gherzi *et al.*, 2001; Van Dijk, Schilders *et al.*, 2007; Shcherbik, Wang *et al.*, 2010). Para verificarmos a expressão do complexo do exossoma em *Leishmania* primeiramente realizamos o “download” na base de dados do NCBI (www.ncbi.nlm.nih.gov) das sequências genômicas dos genes presentes em humanos e em leveduras já previamente descritos e com funcionalidade reconhecida através do banco de dados UniProt (www.uniprot.org). Estas sequências (ver a lista completa na Tabela 6) foram utilizadas como entrada para uma busca dos genes ortólogos presentes em *Leishmania* utilizando a ferramenta “BlastX” no banco de dados do TritypDB (www.tritypdb.org). De posse destas sequências as utilizamos como referências para o mapeamento dos *reads* das cinco bibliotecas transcriptômicas [uma de *L. (L.) amazonensis* e quatro de *L. (V.) braziliensis*]. O mapeamento foi realizado novamente com o software CLC Genomics Workbench (Version 5.0.1). Utilizamos os critérios de 90% de cobertura e 90% de similaridade para todos os mapeamentos.

Todas as análises tiveram o auxílio dos bioinformatas Talles Eduardo Ferreira Maciel (Doutorando do Programa de Pós-Graduação em Bioquímica Agrícola da Universidade Federal de Viçosa) e *M.Sc.* Pedro Marcos Pereira Vidigal (bolsista DTI nível 1 da Universidade Federal de Viçosa).

RESULTADOS

Co-Purificação e Detecção de Diferentes Subunidades de rRNAs Entre os mRNAs poli(A)⁺ de *Leishmania*

A fim de estudar as regiões transcritas em *L. (V.) braziliensis* e *L. (L.) amazonensis* em escala transcriptômica, usamos a abordagem de RNA-Seq (454/Roche) e a estratégia oligo(dT) para isolar as moléculas de mRNA poli(A)⁺ (Wang, Gerstein *et al.*, 2009; Kumar e Blaxter, 2010). Então, os RNAs poli(A)⁺ foram purificados baseados na cauda 3'-poli(A)⁺ ubíqua dos mRNAs eucarióticos usando “beads” magnéticos oligo(dT)-ligados (veja a seção materiais e métodos). No ponto inicial do desenvolvimento do projeto de pesquisa o principal objetivo era avaliar o transcriptoma diferencial em larga escala de duas cepas de *L. (V.) braziliensis* com características polares quanto às suas capacidades infectivas e de virulência tanto em humanos quanto em modelo murino (C57BL/6) para entender melhor os mecanismos de infecção deste parasito (Figura 1).

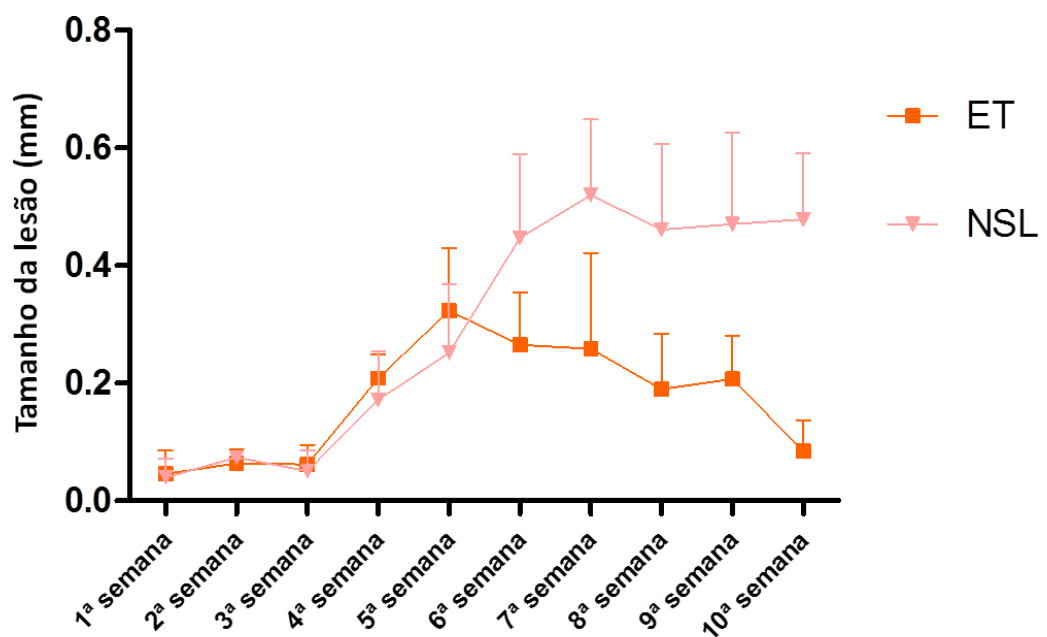


Figura 1. Avaliação da virulência dos isolados ET e NSL de *L. (V.) braziliensis* em modelo murino (C57BL/6). Este resultado mostra que a infecção pelo isolado menos virulento (ET) é controlada pelo hospedeiro, enquanto para o isolado mais virulento (NSL) observa-se a não diminuição da lesão em relação ao tempo de infecção em camundongos.

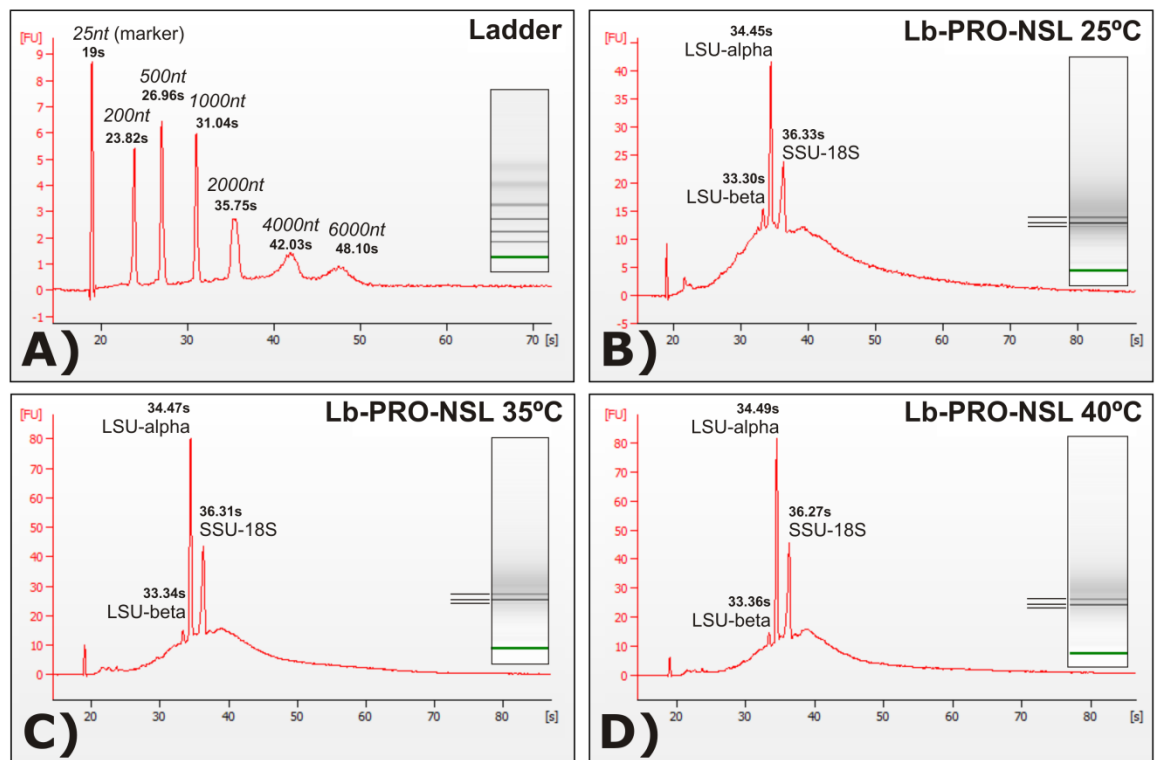
De modo inesperado, no primeiro passo da construção das bibliotecas transcriptômicas, observamos um persistente alto conteúdo de rRNAs nas amostras poli(A)⁺ purificadas. Como podemos ver nos perfis de eletroforese dos RNA poli(A)⁺ purificados de *L. (V.) braziliensis* (promastigotas procíclicos do isolado NSL) que três grandes picos são evidentes (Figura 2B). Estes picos mais abundantes, muito provavelmente, correspondem às subunidades ribossomais, LSU 28S beta, LSU 28S alfa e SSU 18S, respectivamente, porque os tamanhos calculados destas subunidades do rRNA são similares às predições obtidas de dados genômicos para *L. (L.) major* e outros tripanosomatídeos (El-Sayed, Myler, Blandin *et al.*, 2005; Ivens, Peacock *et al.*, 2005).

As subunidades do rRNA apareceram como bandas e picos de RNA bem definidos nas imagens “gel-like” e nos eletroferogramas produzidos pelo Bioanalyzer e correspondem a 11,8% do RNA total poli(A)⁺ (Tabela 2). Em todas as outras extrações realizadas [isolado ET de *L. (V.) braziliensis* e isolado PH8 de *L. (L.) amazonensis*] nos diferentes estágios evolutivos do parasito, perfis similares foram obtidos (Figura Suplementar S1 na seção Anexos). Isto sugere que os rRNAs poderiam ser poliadenilados na extremidade 3'-terminal, similar aos 3'-poli(A)⁺ canônicos dos mRNAs, ou ainda que pequenas extensões internas ricas em As – oligo(dA) – poderiam estar agindo como “isca” nos rRNAs em relação aos oligo(dT) dos “beads” magnéticos.

Para eliminar esta possibilidade e alcançar uma alta pureza nas amostras de mRNAs poli(A)⁺ aumentamos a temperatura de ligação durante a interação poli(A)-oligo(dT). Elevamos a temperatura de hibridização de 25 °C para 35 °C e 40 °C e avaliamos o conteúdo dos rRNAs nas amostras de mRNAs poli(A)⁺ purificados. A ideia foi verificar se a co-purificação das subunidades do rRNA ocorria devido à presença de curtas extensões internas de poli(A) homopoliméricos. Os resultados mostraram que a mesma quantidade de RNA total poli(A)⁺ foi obtida e, mais importante, a mesma proporção de rRNAs foi mantida (Figura 2, letras C e D; Tabela 2). Este resultado reforça a hipótese de que as subunidades do rRNA co-purificadas são devido à presença de caudas poli(A)⁺.

Apoiando esta conclusão, encontramos na literatura um trabalho prévio mostrando que o precursor da subunidade épsilon do rRNA (pré-rRNA LSU 28Sε),

em duas espécies de *Leishmania*, é poliadenilado (Decuypere, Vandesomepele *et al.*, 2005). A presença dos três picos em nossos dados sugere que outras subunidades do rRNA, além do rRNA LSU 28S ϵ , podem ser poliadeniladas também. Em virtude disto, decidimos sequenciar os mRNAs contendo cerca de 10% de possíveis rRNAs poli(A)⁺ a fim de analisar estas sequências de rRNAs mais profundamente. Com isto, o foco deste trabalho passou a ser a investigação da poliadenilação de RNAs ribossomais em *Leishmania*.



Size of rRNA transcripts: 18S=2.2kb alpha=1.78kb gamma=0.21kb zeta=0.073kb
5.8S=0.28kb beta=1.52kb delta=0.18kb epsilon=0.13kb

Figura 2. Padrão de eletroforese capilar das amostras de mRNAs poli(A)⁺ isoladas por afinidade oligo(dT) mostram a presença de subunidades de rRNAs. (A) Padrão do marcador molecular de RNAs (ladder) de uma corrida no Bioanalyzer. Os tempos de retenção são apresentados em segundos acima de cada pico. (B) Padrão eletroforético de uma amostra rica em poli(A)⁺ purificada de promastigotas de *L. (V.) braziliensis* com “beads” magnéticos conjugados a oligo(dT) à 25 °C (Lb-PRO-NSL 25 °C). (C e D) Perfis eletroforéticos das amostras purificadas em 35 °C ou 40 °C de modo similar como descrito em (B). Os tempos de retenção esperados para as diferentes subunidades de rRNAs, calculados de suas conhecidas massas moleculares, estão indicados em cada perfil.

Tabela 2. Quantificação pelo Bioanalyzer – RNA 6000 Pico (teste de ligação).

Grupos de extrações (quantidade da amostra usada)	% ribossomal	Média da % de ribossomais
Lb_PRO_NSL_25°C_5 ng*	11,9	11,8
Lb_PRO_NSL_25°C_10 ng*	11,7	
Lb_NSL_PRO_35°C_5 ng	11,2	12,05
Lb_NSL_PRO_35°C_10 ng	12,9	
Lb_NSL_PRO_40°C_5 ng	12,2	13,2
Lb_NSL_PRO_40°C_10 ng	14,2	

*As amostras foram incubadas por 10 minutos à 25 °C (de acordo com as instruções do fabricante do kit de purificação).

Análises das Bibliotecas de RNA-Seq: Conjunto de Dados das Sequências de rRNAs em Relação às Bibliotecas de Modo Geral

Para a geração das bibliotecas transcriptômicas, as amostras de RNAs poli(A)⁺ purificadas foram convertidas em cDNAs usando “primers” randômicos de 6-mer durante a transcrição reversa (veja a seção Materiais e Métodos). Cinco bibliotecas transcriptômicas foram geradas e sequenciadas: Lb_ET_MET, Lb_NSL_MET, Lb_ET_PRO, Lb_NSL_PRO, La_PH8_MET; onde Lb indica a espécie *L. (V.) braziliensis*; La indica *L. (L.) amazonensis*; ET e NSL indicam as cepas de menor e maior virulência em *L. (V.) braziliensis*, respectivamente; PH8 indica a cepa em *L. (L.) amazonensis*; PRO indica a forma não infectiva promastigota procíclica e MET indica a forma infectiva promastigota metacíclica. Os dados gerais adquiridos em cada corrida no 454/Roche estão resumidos na Tabela 3. O sequenciamento forneceu um total de 4.552.223 *reads* com tamanho médio de 379,4 nucleotídeos (Tabela 3). As sequências apresentaram alta qualidade, como demonstrado pela alta taxa (mais de 88%) dos *reads* com “Phred Score” maior do que 20 em todas as bibliotecas (Ewing e Green, 1998; Ewing, Hillier *et al.*, 1998).

Tabela 3. Resumo dos dados de sequenciamento das bibliotecas transcriptômicas para *L. (V.) braziliensis* e *L. (L.) amazonensis*. Confirmação do alto conteúdo de rRNAs após a purificação dos RNAs poli(A)⁺ por afinidade com beads magnéticos oligo(dT).

Amostra	HQ Reads	Média do tamanho do read	% Reads ribossomais
Lb_ET_MET	916.667	364	125.231 = 13,66%
Lb_NSL_MET	1.083.651	370	312.160 = 28,81%
Lb_ET_PRO	589.929	398	151.916 = 25,76%
Lb_NSL_PRO	506.551	385	108.095 = 21,35%
La_PH8_MET	1.455.425	380	434.141 = 29,83%
Total	4.552.223	379,4	1.131.543 (média = 24,85%)

A fim de analisar os *reads* de rRNAs, agrupamos todas as sequências usando o software CLC Genomics Workbench (Versão 5.0.1) e mapeamos os “contigs” resultantes contra o “cluster” de genes de rRNAs localizado no cromossomo 27 de *L. (L.) major* (Ivens, Peacock *et al.*, 2005). Em *L. (L.) major* os genes de rRNAs SSU

18S, LSU 5.8S e LSU 28S estão fisicamente dispostos em um único longo “cluster” gênico organizado em “tandem” no cromossomo 27 (Ivens, Peacock *et al.*, 2005).

Considerando todas as bibliotecas reunidas, 1.131.543 *reads* de sequências foram mapeados como transcritos de rRNAs, de acordo com o “cluster” de ribossomais em *L. (L.) major*. Este número corresponde a 24,85% do total de *reads* sequenciados (Tabela 3). O conteúdo de *reads* ribossomais nas cinco bibliotecas analisadas variou entre 13,66% a 29,83% (Tabela 3). O “cluster” de rRNAs no genoma de *L. (L.) major*, localizado no cromossomo 27 (989.640 até 1.060.786 nt), corresponde a uma região de 71.146 nucleotídeos que contém 52 genes (www.tritrypdb.org). Considerando o tamanho médio dos *reads* na Tabela 3 e o tamanho do “cluster” de rRNAs em *L. (L.) major*, os 277.147 *reads* de rRNAs do isolado ET de *L. (V.) braziliensis* correspondem a 106.046.652 nucleotídeos e representam uma cobertura de 1.491 vezes. Os 420.255 *reads* do isolado NSL de *L. (V.) braziliensis* correspondem a 157.115.775 nucleotídeos e representam uma cobertura de 2.124 vezes o “cluster” de rRNAs. Da mesma forma, os 434.141 *reads* de *L. (L.) amazonensis* correspondem a 164.973.580 nucleotídeos e uma cobertura de 2.318,8 vezes o “cluster” de rRNAs.

A sequência consenso de cada um dos 52 genes de rRNAs de *L. (V.) braziliensis* (isolados ET e NSL) e *L. amazonensis* (isolado PH8) foram depositados no GenBank sob os códigos de acesso entre JX030037 – JX030192. A Figura 3 mostra de forma esquemática a análise do “cluster” gênico de RNAs ribossomais (rDNA) em relação aos *reads* referentes aos rRNAs encontrados nas bibliotecas transcriptômicas.

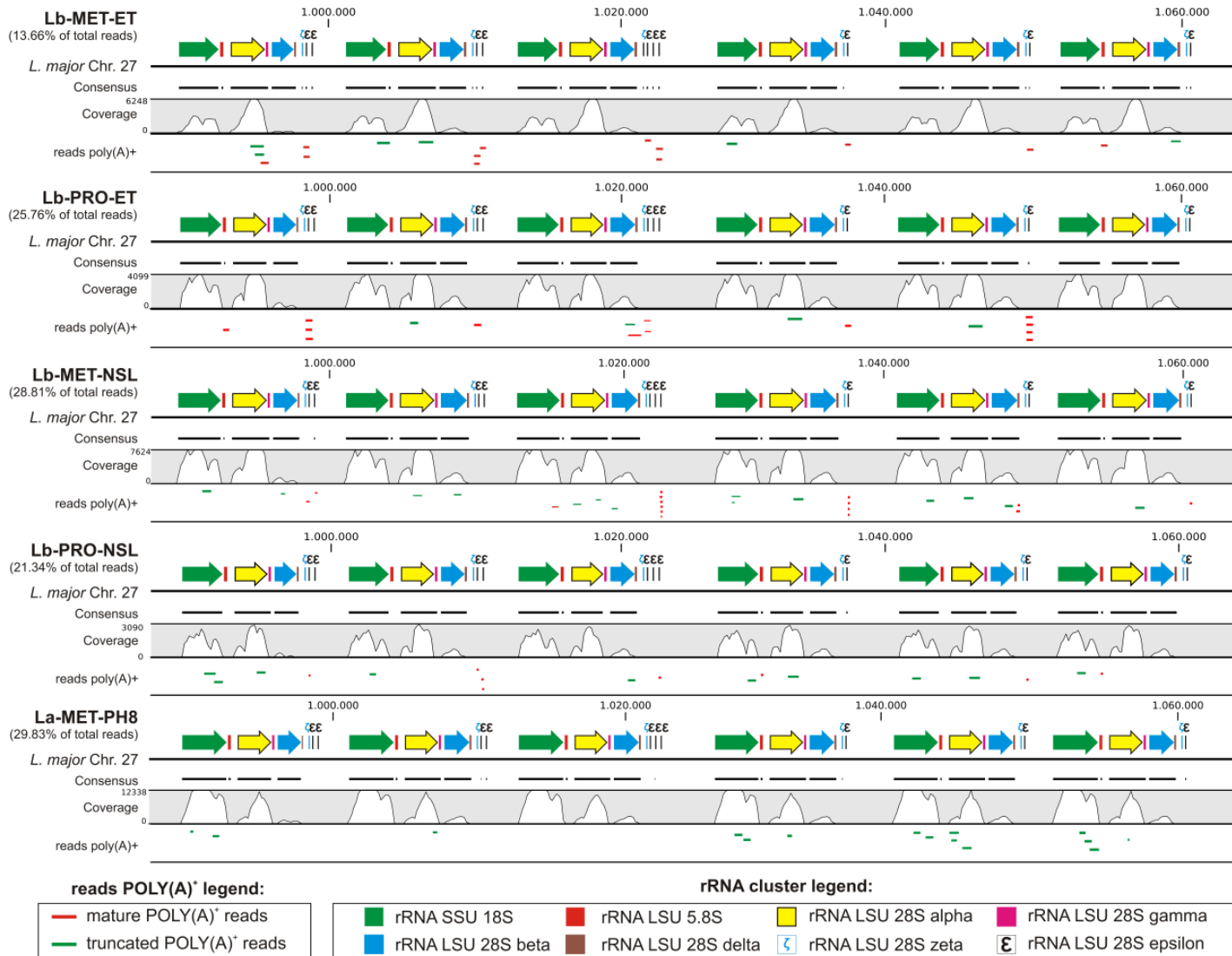


Figura 3. Representação esquemática do mapeamento do “cluster” de RNAs ribossomais de *L. (V.) braziliensis* e *L. (L.) amazonensis*. Dados de transcriptoma por RNA-Seq para o “cluster” de RNAs ribossomais. As sequências de rRNAs foram obtidas das formas metacíclicas (Lb_ET_MET, Lb_NSL_MET) e procíclicas (Lb_ET_PRO, Lb_NSL_PRO) de *L. (V.) braziliensis* e das formas metacíclicas (La_PH8_MET) de *L. (L.) amazonensis* por pirosequenciamento 454/Roche. As sequências foram montadas com o software CLC Genomics Workbench (Version 5.0.1) usando o “cluster” de rRNAs de *L. (L.) major* como “template”. Cada um dos genes de rRNAs sequenciados em ambas as subunidades [maior (do inglês, LSU); e menor (do inglês, SSU)] estão representados por setas grossas e por traços [para os genes pequenos da subunidade maior (LSU)] em cores diferentes. As subunidades precursoras e maduras em *Leishmania* spp. mostraram caudas poli(A)⁺ como indicado pelos traços marcados com “poly(A)⁺reads”: as barras vermelhas horizontais representam os *reads* mapeando a extremidade 3'-terminal dos rRNAs maduros e contendo caudas ricas em poli(A); as barras verdes horizontais representam os *reads* com caudas ricas em poli(A) mapeando em posições internas dos rRNAs truncados. A proporção do conteúdo de rRNAs em relação ao número total de *reads* em cada biblioteca está representado entre parênteses próximo aos nomes das espécies. Note que o número de *reads* poliadenilados de rRNAs que foram detectados e o número de subunidades que foram poliadeniladas foram diferentes entre as espécies e entre as cepas de uma mesma espécie.

Poliadenilação de Moléculas de rRNAs em *L. (V.) braziliensis* e *L. (L.) amazonensis*

Considerando a alta proporção de sequências de rRNAs detectadas nas bibliotecas de RNA poli(A)⁺ em *Leishmania*, decidimos investigar o perfil de poliadenilação nestas sequências mais precisamente. Analisando as extremidades 3'-terminal dos *reads* derivados das subunidades do rRNA, encontramos que algumas delas apresentaram caudas poli(A)⁺ homo- ou heteropoliméricas (Tabela 4; Texto Suplementar S1 na seção Anexos). Os resultados de RNA-Seq sugeriram que os eventos de poliadenilação não são exclusivos da extremidade 3'-matura do pré-rRNA da subunidade épsilon do rRNA LSU 28S como previamente evidenciado (Decuyper, Vandesomepele *et al.*, 2005), mas podem também ocorrer em outras subunidades, tais como no SSU 18S, no LSU 5.8S e em outras subunidades do rRNA 28S (Figuras 2 e 3; Tabela 4). Observamos a presença de caudas poli(A)⁺ na extremidade 3'-matura das subunidades dos rRNAs (linhas vermelhas na Figura 3) e nos rRNAs pré-processados, como evidenciado pelos *reads* contendo uma junção entre as subunidades que não foram completamente processadas (um exemplo está mostrado na Figura 4).

Observamos também que muitas sequências alinharam muito bem com o “cluster”, mas outras desalinham totalmente com as regiões espaçadoras intergênicas (Figura 3). Estes dados sugerem que a maioria das moléculas de rRNAs purificadas durante as construções das bibliotecas foram realmente processadas liberando as subunidades separadas dos rRNAs. Entretanto, não podemos eliminar a possibilidade de que os espaços intergênicos não sejam conservados entre as espécies, uma vez que o “template” das sequências utilizadas para o alinhamento *in silico* foi o genoma de *L. (L.) major*.

Os distintos padrões das caudas poli(A)⁺ observadas nas sequências dos rRNAs estão sumarizados na Figura 4, como se segue: (A) rRNA maduro com adição de cauda poli(A) na extremidade 3'-terminal; (B) rRNA pré-processado com adição de cauda poli(A) na extremidade 3'-terminal; (C) rRNA com adição de cauda poli(A) interna; (D) rRNA pré-processado com adição de cauda poli(A) interna.

Tabela 4. Moléculas de rRNA SSU 18S, LSU 5.8S e LSU 28S rRNA poliadeniladas mapeadas nos dados de transcriptoma em *Leishmania* spp.

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_SSU_18S LmjF27.rRNA.02	GKLMG1101CH44D	ET/Metacyclic	2204	2074	A ₂₃ CA ₇ NA ₆ A₆CA₁₂
Lb_rRNA_SSU_18S LmjF27.rRNA.04	GKLMG1102FZA86	ET/Metacyclic	2204	911*	G ₂ A ₉ G ₂ CAG
Lb_rRNA_LSU_5.8S LmjF27.rRNA.12	GKLMG1102GJ61H	ET/Metacyclic	283	284	A ₆ CA ₉ A₂₂NA₇
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.13	GKLMG1101C37RG	ET/Metacyclic	1780	1580*	A ₈ NA ₅ NA ₁₆ N
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.13	GKLMG1101DYMLN	ET/Metacyclic	1780	1526*	ACTA ₁₆ NA₈NA₁₀NA₉NA₉CA₈
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.13	GKLMG1102IDFEX	ET/Metacyclic	1780	1781	C ₂ TC ₂ T ₂ CTGAGTGA ₂ GA ₃ CA ₁₄ A₂NA₇GA₃ CGTCGT
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.14	GKLMG1102IDFEX	ET/Metacyclic	1780	1649	A ₄ GA ₃ GA ₁₀ CA ₂ GA ₃ G
Lb_rRNA_LSU_28S_beta + Lb_rRNA_LSU_delta LmjF27.rRNA.30 + LmjF27.rRNA36	GKLMG1102F8T5P	ET/Metacyclic	1525 + 183	122*	ACA ₂ CA ₁₃ NA₈NA₁₁NA₁₀NA₇T

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GKLMG1102HUX3F	ET/Metacyclic	129	128	A ₁₁ NA ₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GKLMG1102GZJPE	ET/Metacyclic	129	127	ACA ₉ A ₄₃ NA ₅ N
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.45	GKLMG1102FK9WN	ET/Metacyclic	129	128	A ₉ NA ₈ NA ₉ NA ₁₇ NA ₈ NA ₈ NA ₆ GA ₃ G
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.45	GKLMG1102IXJB6	ET/Metacyclic	129	128	A ₉
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.46	GKLMG1102HMWKE	ET/Metacyclic	129	127	ACA ₂₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.47	GKLMG1101DOYL7	ET/Metacyclic	129	127	ACA ₇ NA ₁₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GKLMG1101CMLJZ	ET/Metacyclic	129	126	ACA ₈ A ₃₅ NA ₁₇ NA ₈ NA ₈ NA ₈ NA ₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GKLMG1102IQS38	ET/Metacyclic	129	128	A ₇
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GKLMG1102IR6X9	ET/Metacyclic	129	128	A ₆ NA ₁₆

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.51	GKLMG1102H6IYU	ET/Metacyclic	129	127	ACA ₁₉ NA ₈ NA ₁₈ NA ₈ NA ₂₁ N
Lb_rRNA_SSU_18S LmjF27.rRNA.01	GPST6G001A28QK	NSL/Metacyclic	2204	1719*	GATGA ₁₀ C
Lb_rRNA_SSU_18S LmjF27.rRNA.03	GPST6G001ASD22	NSL/Metacyclic	2204	2197	A ₁₄ G
Lb_rRNA_SSU_18S LmjF27.rRNA.04	GPST6G001ASH89	NSL/Metacyclic	2204	1249*	TAGCA ₁₃
Lb_rRNA_SSU_18S LmjF27.rRNA.05	GPST6G002GNP5P	NSL/Metacyclic	2204	1882*	A ₇ GCA ₃₆
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.14	GPST6G002F9X2K	NSL/Metacyclic	1780	1048*	CAT ₃ AG ₂ TA ₃ GTA ₁₆ NA ₁₂ NA ₁₁ NA ₁₀ GA ₇ G ₄ A ₃ G ₃
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.15	GPST6G001COWY2	NSL/Metacyclic	1780	490*	CA ₂ GA ₃ CAGAGA ₁₀
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.15	GPST6G002I4OM4	NSL/Metacyclic	1780	1553*	AGA ₁₁
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.16	GPST6G002GRQI3	NSL/Metacyclic	1780	1556*	A ₁₃ A ₉ NA ₅₁

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GPST6G002G0NZ5	NSL/Metacyclic	1780	1153*	A ₅ TA ₂ CGTA
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.18	GPST6G002GQO2Y	NSL/Metacyclic	1780	1556*	A ₂₀ A ₁₄ NA ₁₉ NA ₉ NA ₈ NA ₇ NA ₆ NA ₄
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.25	GPST6G001BBTQP	NSL/Metacyclic	1525	628*	A ₁₀ A ₅₁ NA ₆
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.26	GPST6G002H92OB	NSL/Metacyclic	1525	1220*	A ₉ NA ₉ GT
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.27	GPST6G001C06JQ	NSL/Metacyclic	1525	261*	AGA ₁₅ CA ₁₃
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.29	GPST6G002IIJK1	NSL/Metacyclic	1525	1219*	A ₈ NA ₈ N
Lb_rRNA_LSU_28S_delta LmjF27.rRNA.35	GPST6G001A9OGW	NSL/Metacyclic	183	184	CA ₃ CA ₂ CA ₁₇ A ₇ NA ₁₁ NA ₁₀ NA ₉ NA ₈ NA ₅
Lb_rRNA_LSU_28S_delta LmjF27.rRNA.35	GPST6G002IL7M1	NSL/Metacyclic	183	184	CA ₃ CA ₂ CTA ₃ T ₃ A ₂₇ A ₇ NA ₇ NA ₈ NA ₉ NA ₇
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GPST6G001D57MS	NSL/Metacyclic	129	129	A ₁₇ NA ₂₁ N

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.44	GPST6G002F9KTJ	NSL/Metacyclic	129	125	CA ₂₂
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G002JT230	NSL/Metacyclic	129	124	A ₁₄ NA ₁₅ A₃₂NA₈NA₇NA₆NA₃G
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G002G4PTN	NSL/Metacyclic	129	129	A ₁₂ A₇₀NA₆NA₃G
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G001B1LX8	NSL/Metacyclic	129	129	A ₁₆ A₁₄NA₆GA₃GA
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G002G8BFU	NSL/Metacyclic	129	126	CA ₁₁ A₈NA₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G002I4NH2	NSL/Metacyclic	129	127	A ₂ CA ₁₆ A₆₅N
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPST6G002FMXGZ	NSL/Metacyclic	129	129	A ₂₉ A₃₅NA₃N
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GPST6G002GZB77	NSL/Metacyclic	129	126	ACA ₁₁ A₁₂NA₁₂NA₁₀
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GPST6G002I3B5R	NSL/Metacyclic	129	129	A ₁₆ A₅₄NA₅NA₃NA₂N

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GPST6G002HCSP1	NSL/Metacyclic	129	126	CA ₁₂ A ₁₀ NA ₁₀ NA ₈ GA ₅ GA ₂ G
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GPST6G002JKMMI	NSL/Metacyclic	129	129	A ₁₈ A ₃₉ NA ₈ NA ₇ NA ₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.52	GPST6G001A90XR	NSL/Metacyclic	129	129	A ₂₀ NA ₉ NA ₈ NA ₅ G
Lb_rRNA_LSU_5.8S LmjF27.rRNA.07	GPUPS7E01B33ZO	ET/Procylic	283	284	A ₆ CA ₁₂ NA ₃₄ NA ₁₈ NA ₇ NA ₅ N
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.14	GPUPS7E01DQJYH	ET/Procylic	1780	815*	TCTC ₂ T ₂ C ₂ TCGTGCACGTCTGC ₂ GCGCAT GCACACGCGCGCGT ₂ GAC ₂ GTA ₁₈
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.16	GPUPS7E01DBYAX	ET/Procylic	1780	1555*	A ₁₆
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GPUPS7E01AJ04T	ET/Procylic	1780	1544*	AGA ₁₆ NA ₁₁ NA ₆ CA ₅ NA ₅ NA ₇
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.27	GPUPS7E01DTIZS	ET/Procylic	1525	1316*	ACTA ₁₅ NA ₂ A ₄₂ NA ₁₀ NA ₉ NA ₈ NA ₈ NA ₇ NA ₆

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_beta + Lb_rRNA_LSU_delta LmjF27.rRNA.27 +LmjF27.rRNA.33	GPUPS7E01CY0WT	ET/Procyclic	1525 + 183	184	CA ₃ CA ₂ CTA ₂ T ₃ A ₁₅ NA₁₄NA₁₃NA₁₂NA₇C
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GPUPS7E01DGSP1	ET/Procyclic	129	125	CA ₁₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GPUPS7E01BTIQD	ET/Procyclic	129	129	A ₁₁ A₁₇NA₈TA₄TA₂
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GPUPS7E01BT560	ET/Procyclic	129	129	A ₂₆ NA ₁₂ NA₁₁NA₉NA₉NA₇N
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.45	GPUPS7E01CGLVD	ET/Procyclic	129	129	A ₄ CA ₉ A₂₉NA₈NA₇NA₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.47	GPUPS7E01DL1L5	ET/Procyclic	129	127	ACA ₁₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.47	GPUPS7E01EELT5	ET/Procyclic	129	129	A ₉ NA ₁₉ NA ₁₀ NA ₉ NA ₇ CT

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.50	GPUPS7E01C875V	ET/Procyclic	129	128	A ₁₅ A ₃₆ NA ₇
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.51	GPUPS7E01AGE78	ET/Procyclic	129	126	CA ₁₁
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.51	GPUPS7E01B91TY	ET/Procyclic	129	129	A ₁₂ A ₅₃ NA ₇ NA ₆ NA ₄
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.51	GPUPS7E01A3X1U	ET/Procyclic	129	129	A ₁₆ A ₄₂
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.51	GPUPS7E01DO5CX	ET/Procyclic	129	127	ACA ₉ A ₁₅ NA ₈ TA ₆ TA ₃
Lb_rRNA_SSU_18S LmjF27.rRNA.01	GPUPS7E02HR03W	NSL/Procyclic	2204	2051	ACA ₁₄
Lb_rRNA_SSU_18S LmjF27.rRNA.01	GPUPS7E02GGL7T	NSL/Procyclic	2204	1459*	TG ₂ CT ₂ C ₂ GCGT ₂ GCTCT ₂ CT ₄ CTC ₃ TCT ₂ GT G ₃ CTG ₂ ATCAG ₂ TGCGCGCG ₂ CGTAGTGC G ₂ CGTAGTGCTGTGTGTGTGTCTCTCTCT ₂ GTGCAG ₃ ACT ₂ GTAG ₂ T ₂ AGTAGCA ₃ CTC AT ₂ C ₄ TNTCGTAGCG ₂ A ₃ T ₂ A ₁₇ NA ₁₄ NA ₁₅ N A ₁₃ NA ₁₂ NA ₈ TA ₄
Lb_rRNA_SSU_18S LmjF27.rRNA.02	GPUPS7E02FFUJJ	NSL/Procyclic	2204	1486*	A ₁₀

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_SSU_18S LmjF27.rRNA.04	GPUPS7E02JJELS	NSL/Procyclic	2204	2052	ACA ₁₅
Lb_rRNA_SSU_18S LmjF27.rRNA.05	GPUPS7E02GC8AZ	NSL/Procyclic	2204	986*	TCT ₂ GCTGCGTGTG ₂ A ₂ GAG ₅ AG ₃ CG ₅ AG ₂ AG ₃ AGT ₂ G ₄ CGTAG ₂ TGCC ₂ TACA ₉ CGA ₁₂ NA₁₄NA₁₅
Lb_rRNA_SSU_18S LmjF27.rRNA.06	GPUPS7E02GPDEF	NSL/Procyclic	2204	1541*	A ₁₀ CAT
Lb_rRNA_LSU_5.8S LmjF27.rRNA.10	GPUPS7E02HKAKA	NSL/Procyclic	283	284	A ₆ CA ₁₁ NA ₁₃ NA ₁₁ A₂₇NA₅N
Lb_rRNA_LSU_5.8S LmjF27.rRNA.12	GPUPS7E02GP34K	NSL/Procyclic	283	284	A ₆ CA ₈ ANA₃₆NA₅N
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.13	GPUPS7E02IJPKY	NSL/Procyclic	1780	1548*	A ₁₃ NA₃₁NA₁₉NA₁₆NA₅NA₅N
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.16	GPUPS7E02JBMN5	NSL/Procyclic	1780	1372*	AGA ₁₅ NA₁₀NA₁₀CA₆NA₈NA₉NA₉NA₁₀NA₈NA₅
Lb_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GPUPS7E02GRRJQ	NSL/Procyclic	1780	1552*	ACA ₁₈

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
Lb_rRNA_LSU_28S_beta LmjF27.rRNA.27	GPUPS7E02HOF3	NSL/Procyclic	1525	1437	TATGA ₂ CA ₃ TCATAC ₂ AGAG ₂ TGTCAGA ₂ TA T ₂ AC ₂ ACAG ₃ ATA ₂ CTG ₂ CT ₂ GTG ₂ CG ₂ C ₂ A ₂ GCGT ₂ CATAGCGACGTCGCT ₅ GATC ₂ T ₂ CG ATGTCG ₂ CTCT ₂ C ₂ TA ₂ C ₂ TAGCGC ₂ GCAGA ₂ GACGCTA ₂ G ₃ T ₂ G ₂ AT ₂ GT ₂ CAC ₃ ACTGACA G ₃ A ₂ CGTGAGCTG ₃ T ₃ AGAC ₂ GTCGTGAGA CAG ₂ T ₂ G ₂ T ₄ AC ₃ TACT ₂ AGCTAG ₂ A ₁₆ NA₁₀N A₂₂NA₁₀NA₁₀NA₈NA₉NA₈NA₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.43	GPUPS7E02IUMQ6	NSL/Procyclic	129	129	ACA ₂₀ NA₈CA₃CACAC
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.45	GPUPS7E02FPUSB	NSL/Procyclic	129	128	A ₃₅ NA₁₃GA₇TA₃G
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.46	GPUPS7E02GPH29	NSL/Procyclic	129	129	A ₂₂ A₅
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.46	GPUPS7E02GGFXE	NSL/Procyclic	129	129	A ₇ A₄₁NA₈NA₆
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPUPS7E02GNWVT	NSL/Procyclic	129	129	A ₂ A₂₇NA₁₅NA₁₃NA₁₁NA₁₁NA₉NA₇CA₄
Lb_rRNA_LSU_28S_epsilon LmjF27.rRNA.49	GPUPS7E02F0JN1	NSL/Procyclic	129	128	A ₂₀ NA₉NA₁₈NA₁₀NA₉NA₈NA₇NA₅N

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
La_rRNA_SSU_18S LmjF27.rRNA.01	GWOOT3C01COX8Y	PH8/Metacyclic	2204	487*	A ₂ T ₂ G ₂ T ₂ C ₂ A ₉ C ₂ A ₉ C ₂ G ₃ A ₃
La_rRNA_SSU_18S LmjF27.rRNA.01	GWOOT3C01DH9K0	PH8/Metacyclic	2204	1711*	ATCTGCT ₂ C ₂ ATC ₂ GTGACA ₂ GACATGAG ₂ A ₃ C ₂ GA ₂ GTA ₂ C ₂ A ₂ CGGA ₂ GA ₄ CACGTGTG AG ₃ TGA ₄ GAG ₂ A ₂ GACACGAGCA ₂ CATGAT CACAGCACAGCAC ₂ A ₃ CGCGA ₂ G ₂ A ₅ GA ₁₃ C
La_rRNA_SSU_18S LmjF27.rRNA.04	GWOOT3C01BPZKL	PH8/Metacyclic	2204	1787*	GTA ₂ TA ₂ TATATATATATA ₂ GGACACA ₂ GC GCACAG ₂ CGA ₃ CAGACA ₂ CACACGC ₂ G ₂ GA ₃ NA ₄ C ₂ ACACACACACATACATACA ₂ CA ₈ G ₂ A ₃ CA ₅ CA ₄ CTA ₉ CA ₃ CGCTACGCATGTA ₂ CG TCA ₅ C ₂ GA ₂
Lb_rRNA_SSU_18S LmjF27.rRNA.04	GWOOT3C02JLGXG	PH8/Metacyclic	2204	1473*	ACACGCA ₂ CA ₄ CAGCA ₂ GA ₈
La_rRNA_SSU_18S LmjF27.rRNA.05	GWOOT3C01BKANS	PH8/Metacyclic	2204	1787*	GTA ₂ TA ₂ TATATATATATA ₂ G ₂ ACACA ₂ GCG CACAG ₂ CGA ₄ CAGACA ₂ CACACGC ₂ G ₂ GA ₂ GA ₄ C ₂ ACACACACACATACATACA ₉ G ₂ A 4CA₆CA₅CTA₁₀CA₃CGACTACGA
La_rRNA_SSU_18S LmjF27.rRNA.05	GWOOT3C01EWSO4	PH8/Metacyclic	2204	1156*	TGAGATACATAGA ₂ G ₂ GA ₂ GCA ₃ GA ₃ CGA ₂ CGCACGCACGCAGATACACA ₂ GA ₂ TGGA ₄ C ₂ ACA ₃ G ₂ A ₂ GA ₂ CA ₃ CAGACAGAGAGAGA GAGATGTATCGCGTGCCTG ₂ AG ₂ TG ₂ A ₃ GT A ₃ CGA ₉ CGA ₇ GA ₂ G ₂

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
La_rRNA_SSU_18S LmjF27.rRNA.06	GW00T3C01B73UY	PH8/Metacyclic	2204	2160	G ₄ AG ₂ CTG ₂ AGA ₂ CGAGAGTG ₂ CAGCACTG A ₂ GAGT ₄ ACGTGGTGTG ₃ CA ₂ CA ₂ TGGA ₆ G A₇G₂A₂TA₂CA₆CGACA₃C₂A₂TA₈CGA₃GTA GACG₂
La_rRNA_SSU_18S LmjF27.rRNA.06	GW00T3C01E07AP	PH8/Metacyclic	2204	1787*	GTA ₂ TA ₂ TATATATATATA ₂ G ₂ ACACA ₂ GCG CACAG ₂ CGA ₃ CAGACA ₂ CACACGC ₂ G ₂ GA ₂ GA ₄ C ₂ ACACACACACATACATACACA ₉ G ₂ A 3CA₅CA₄CTA₉CA₃CGACT
La_rRNA_SSU_18S LmjF27.rRNA.06	GW00T3C01EIGMS	PH8/Metacyclic	2204	1557*	C ₃ AC ₃ TG ₂ ATCGTATGCGTATATAGAT ₂ A ₈ C AGA ₂ CA ₃ CA ₂ C
La_rRNA_LSU_28S_alpha LmjF27.rRNA.14	GW00T3C01A0IV6	PH8/Metacyclic	1780	1744	ATAGA ₇ GAG ₂ AGATGAGCAGATG ₂ AG ₂ A ₂ GACG ₂ TCA ₂ TGTAGCGCAGAT ₂ A ₂ CTGCTC GA ₃ CTCA ₃ CGA ₂ CGACA ₈ CAGAGA ₂
La_rRNA_LSU_28S_alpha LmjF27.rRNA.16	GW00T3C01B2EN0	PH8/Metacyclic	1780	1077*	GA ₃₀ T ₂ CGAC ₂ TAT ₂ CTCTA ₃ CT ₃ A ₃ TGTGCA ₃ CA ₃ C ₂ GTGC ₂ T ₂ AGC ₂ A ₂ CTGCA ₂ G ₂ AC ₂ G ₂ A ₃ GA ₄ GA ₂
La_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GW00T3C01CL3IZ	PH8/Metacyclic	1780	379*	G ₂ CGTCG ₂ A ₄ GCA ₂ TCTAGA ₃ CATA ₂ C ₂ A ₂ C ACACGAGCACGCACGCGCGCGCGCGCA CGCAGCATAGACA ₂ GCA ₂ CACGACGAGA ₂ CA ₃ CA ₂ TA ₃ CA ₁₁ GA₄GA₂CA₂GAC₂A₂G₂A 3CGA₃T

Nome do gene Gene template	Identidade do <i>Read</i>	Cepa/Forma evolutiva	Tamanho total (nt)	Sítio poli(A) ⁺ (nt)	Tamanho e conteúdo da cauda**
La_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GWO0T3C01EN1ST	PH8/Metacyclic	1780	306*	GAG ₂ AT ₂ GAGCTGTAG ₂ GA ₂ CG ₂ TG ₂ T ₃ CA GCAGAGTGCAG ₂ ATGAGA ₃ G ₂ TACAGC ₂ GT ₂ GCTC ₂ GTGCATGTGAGCGT ₄ CAGAC GTA ₈ CA ₆ GA ₄ CAC ₂ T
La_rRNA_LSU_28S_alpha LmjF27.rRNA.17	GWO0T3C02JPXBS	PH8/Metacyclic	1780	1113*	TGCA ₂ GA ₂ CA ₈ CTCGA ₅ CA ₆ CGA ₂
La_rRNA_LSU_28S_alpha LmjF27.rRNA.18	GWO0T3C01C02EM	PH8/Metacyclic	1780	1053*	A ₆ TGAT ₂ AGAG

* Para as caudas poli(A)⁺ consideradas como internas, como critério de análise, estavam localizadas ao menos 200 nucleotídeos acima da extremidade 3'-terminal dos genes para as subunidades 18S, 28S alfa e 28S beta. Para os outros genes, 50 bases acima da extremidade 3'-terminal das subunidades foram consideradas. **As sequências das bases tem "score" de qualidade > 10. Os resíduos marcados em negrito apresentam um "score" de qualidade ≤ 10.

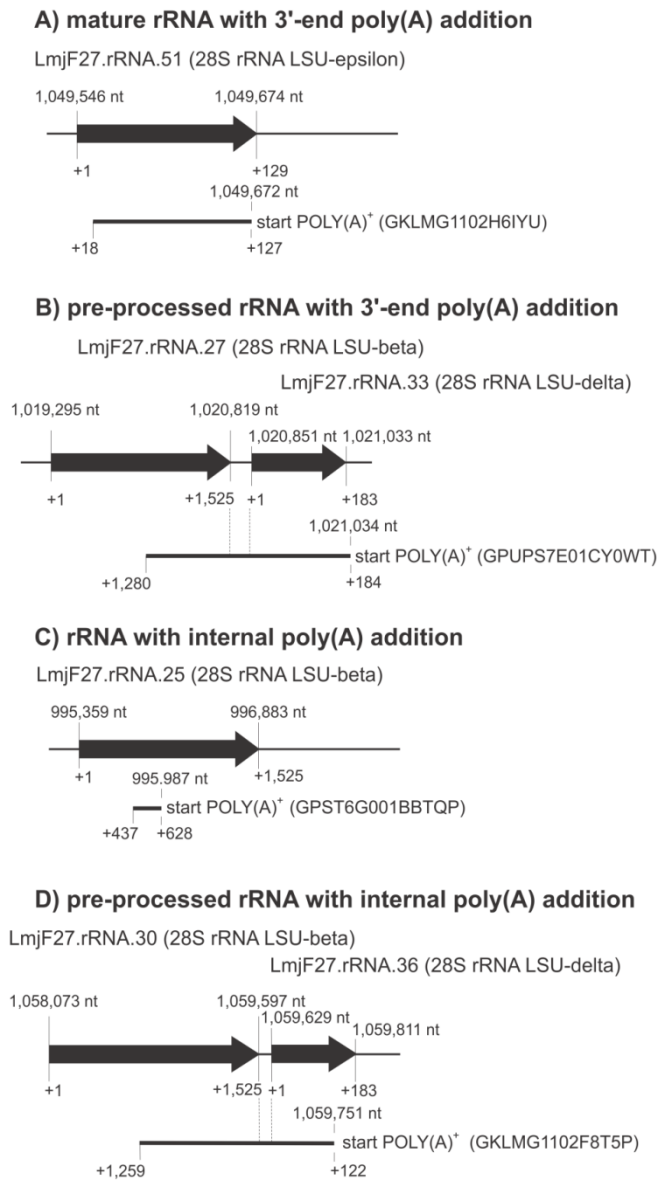


Figura 4. Exemplos de fragmentos de rRNAs com caudas ricas em poli(A) em *Leishmania*. (A) subunidade épsilon madura do rRNA 28S do genoma de *L. (L.) major* está esquematicamente representada com uma seta; o *read* gerado por RNA-Seq de *L. (V.) braziliensis* que mapeou neste rRNA está mostrado logo abaixo, com as coordenadas do mapeamento indicadas. O ponto de início do poli(A) na extremidade 3'-madura do *read* RNA-Seq está indicado. A identidade da sequência (sequence_ID) está mostrada próxima ao *read* esquemático. (B) “cluster” do DNA genômico do rRNA 28S de *L. (L.) major* com duas subunidades (28S LSU-beta e 28S LSU-delta) está representado por duas setas grossas. Um *read* RNA-Seq mapeando nestas duas subunidades e com uma cauda rica em poli(A) na extremidade 3'-terminal da subunidade LSU-delta é mostrado logo abaixo. A identidade da sequência é mostrada à direita do esquema do *read* RNA-Seq. (C) subunidade LSU-beta madura do rRNA 28S do genoma de *L. (L.) major* está esquematicamente representada com uma seta; o *read* RNA-Seq de *L. (V.) braziliensis* que mapeou numa região interna deste rRNA está representado logo abaixo, com as coordenadas do mapeamento indicadas. A sequência poli(A) na extremidade 3'-terminal do *read* RNA-Seq está indicada. A identidade da sequência está mostrada próxima ao *read* esquemático. (D) “cluster” do DNA genômico do rRNA 28S de *L. (L.) major* com duas subunidades (28S LSU-beta e 28S LSU-delta) está representado por duas setas grossas. Um *read* RNA-Seq mapeando estas duas subunidades e com uma cauda rica em poli(A) iniciando num ponto interno na região codificante da subunidade 28S LSU-delta está mostrado logo abaixo. A identidade da sequência está mostrada à direita do esquema do *read* RNA-Seq.

Poliadenilações Internas em rRNAs e mRNAs Sugerem um Novo Processamento Pós-Transcricional de RNAs em *Leishmania*

A mais surpreendente descoberta neste trabalho foi a observação das poliadenilações em posições internas dos rRNAs e mRNAs. Analisando as posições das extensões poli(A)⁺ nas sequências dos rRNAs encontramos *reads* nos quais a cauda poli(A)⁺ interrompeu as regiões correspondentes às sequências maduras dos rRNAs (Figura 3, linha verde). Isto pode ser mais bem observado na Tabela 4 onde comparamos o tamanho total das sequências codificantes com a posição onde a cauda poli(A)⁺ se iniciou (Tabela 4, linhas destacadas pelo símbolo *). Estes dados evidenciaram que estes *reads* são rRNAs truncados que poderiam não ser usados para construir os ribossomos funcionais do complexo polissômico, sugerindo um novo tipo de processamento de RNA que poderia estar envolvido com a estabilidade dos rRNAs.

Este fenômeno observado em *Leishmania* parece não ser restrito aos rRNAs: as mesmas extensões poli(A)⁺ foram encontradas em *reads* correspondentes a mRNAs que codificam distintas proteínas (Tabela 5; Texto Suplementar S2 na seção Anexos). Considerando que estas caudas poli(A)⁺ interrompem as regiões codificadoras, este grupo de mRNAs truncados não deve estar sendo usado nos eventos de tradução.

Neste ponto surgiram novas questões: é possível que estes RNAs truncados poliadenilados sejam alvos de alguma maquinaria de degradação? Se este for o caso, eles são parte de uma via (ainda não descrita) de decaimento de RNAs guiadas por poliadenilação em *Leishmania*? Para responder a estas perguntas, revisamos a literatura sobre mecanismos de degradação de RNAs. Assim, verificamos que em outros organismos, os RNAs truncados poliadenilados são degradados pelo complexo do exossomapoli(A)⁺-direcionados (Fleischmann, Liu *et al.*, 2004; Kuai, Fang *et al.*, 2004; Deutscher, 2006; Slomovic, Laufer *et al.*, 2006; Slomovic, Fremder *et al.*, 2010). Adicionalmente, nos demais organismos já analisados, caudas homo- ou heteropoliméricas já foram identificadas e são adicionadas em sítios variados (Kuai, Fang *et al.*, 2004; Kao e Read, 2005; Slomovic, Fremder *et al.*, 2010), e o mesmo padrão foi observado nos dados de transcriptoma em *Leishmania* (Figura 3; Tabelas 4 e 5). Isto reforça a hipótese de que em *Leishmania* os RNAs

truncados poliadenilados sejam direcionados para degradação e abrem a possibilidade de que os parasitos do gênero *Leishmania* possam ser capazes de realizar uma via de degradação similar.

Tabela 5. Moléculas de mRNAs truncadas poliadeniladas mapeadas através das análises do dados do transcriptoma para os genes codificantes de proteínas em *L. (V.) braziliensis*.

Nome do gene	Read ID	Cepa/Forma evolutiva	Número de reads	Tamanho da sequência cosenso	Tamanho da sequência referência	Sítio poli(A) ⁺ interno (nt)	Média de cobertura (%)	Tamanho da cauda poli(A) ⁺ e conteúdo **
LbrM.11.0550_40S ribosomal protein S21 putative	GKLMG1101B8T4O	ET/Metacyclic	299	498	498	463	100	A ₃ GCTCA ₃ TGA ₄ GA ₆ G ₃ TG A ₄ CGAGA ₂ GAC ₃ ACAGC AGTA ₄ GAG ₂ AG ₃ CTGCT G ₂ T ₃ C ₃ TCTCT ₂ GACTAGT G ₂ TAGA ₇ GAG ₄ C ₂ GTCGT AT ₄ CT ₃ AC ₂ T ₃ A ₂ CT ₂ A ₁₉ C
LbrM.25.0590_eukaryotic initiation factor 5a putative	GKLMG1101DBLE9	ET/Metacyclic	126	501	501	217*	100	A ₂
LbrM.26.0120_adenine phosphoribosyltransferase	GKLMG1102G60U9	ET/Metacyclic	9	576	651	577	88,48	A ₅ CGA ₁₈
LbrM.34.1800_60S ribosomal protein L5 putative	GKLMG1102INSWN	ET/Metacyclic	750	918	918	365*	100	A ₂ CGA ₁₃
LbrM.29.2350_kinesin putative	GPUPS7E02F1JNG	NSL/Procylic	86	1557	1557	1093*	100	A ₁₀
LbrM.20.1090_kinesin putative	GKLMG1102F1B2V	ET/Metacyclic	286	2858	2862	1245*	99,86	A ₁₀
LbrM.03.0340_protein kinase putative	GPUPS7E01EZ55B	ET/Procylic	62	4295	4527	1457*	94,88	A ₁₄

Para as caudas poli(A)⁺ consideradas como internas, como critério de análise, estavam localizadas ao menos 200 nucleotídeos acima da extremidade 3'-terminal dos genes. ** As sequências das bases tem score de qualidade > 10.

Transcritos de rRNAs Internamente Poliadenilados e não Abundantes são Confirmados por Oligo(dT) RT-PCR

As caudas internas poli(A)⁺ ou ricas em poli(A)⁺ em outros organismos estão associadas ao decaimento de RNAs agindo como marcadores que direcionam os RNAs alvos para vias de degradação, tornando muito difícil a identificação destas moléculas por serem transientes (Kuai, Fang *et al.*, 2004; Slomovic, Fremder *et al.*, 2010). Do mesmo modo, nossos dados de transcriptoma sugeriram a presença deste fenômeno em *Leishmania* somente por um pequeno número de *reads* contendo estes “tags” nas bibliotecas transcriptômicas (Figura 3; Tabelas 4 e 5). Então, o próximo passo foi provar a presença de caudas poli(A)⁺ por RT-PCR direcional e sequenciamento pelo método de Sanger (Sanger, Nicklen *et al.*, 1977). Assim, desenhamos uma estratégia de amplificação usando “primers” 5’ específicos e um “primer” reverso 3’ oligo(dT) não ancorado ligado a um adaptador que serviu como “primer” reverso nas subseqüentes etapas de PCR, como previamente descrito (Slomovic, Portnoy *et al.*, 2008).

A presença de caudas canônicas 3’-poli(A)⁺ e caudas internas poli(A)⁺ ou ricas em poli(A)⁺ foi verificada em moléculas da subunidade beta do rRNA (LSU 28S) e do mRNA do gene *eiF5a2*, ambas utilizadas como modelos de validação. Seguente à transcrição reversa oligo(dT), realizamos as reações de PCR para os genes LSU 28S beta e *eiF5a2*. Os produtos das reações de RT-PCR foram submetidos à eletroforese em gel de agarose e corados com brometo de etídio (Figura 5). Nesta abordagem, esperávamos bandas específicas relacionadas aos “amplicons” contendo caudas canônicas 3’-poli(A)⁺ (668 pb para o gene rRNA 28S beta e 430 pb para o mRNA *eiF5a2*) e um arraste (“smear”) abaixo das bandas canônicas representando os genes truncados com caudas poli(A)⁺, uma vez que estas poliadenilações não são sítio-específicas. Deve ser notado que o rRNA total celular, tanto o precursor (pré-rRNA nuclear) quanto o rRNA maduro (localizado no citoplasma) foram analisados aqui, pois usamos RNA total de extrato celular total. Como as moléculas poliadeniladas foram truncadas em vários sítios ao longo de ambos os genes, e estas extensões poli(A)⁺ foram de diferentes comprimentos, a acumulação não resultou numa única banda de forte intensidade [representando os sítios canônicos 3’-

poli(A)⁺], mas num sinal “borrado” (“smear”), logo abaixo da respectiva banda forte de cada gene (Figura 5).

A detecção destas moléculas truncadas poliadeniladas foi somente possível com um número maior de ciclos saturantes de PCR (20 a 25 ciclos), especialmente para o gene eiF5a2, o que indicou que estas moléculas não são abundantes na célula e, ou, o “turnover” é muito rápido e o processo de degradação é muito eficiente (Figura 5, colunas 9 a 12 no painel da direita). A banda de maior intensidade na Figura 5 (painéis direito e esquerdo) representa a amplificação do produto gênico maduro 3'-poli(A)⁺ do gene eiF5a2 com uma banda de tamanho similar ao esperado de aproximadamente 430 pares de bases, desconsiderando a 3'-UTR (Figura 5, colunas 1 a 4 e 9 a 12). De modo interessante, a amplificação do produto maduro predito do gene LSU 28S beta resultou numa banda forte inferior a 500 pb, abaixo do tamanho esperado de 668 pb (Figura 5, colunas 5 a 8 e 13 a 16).

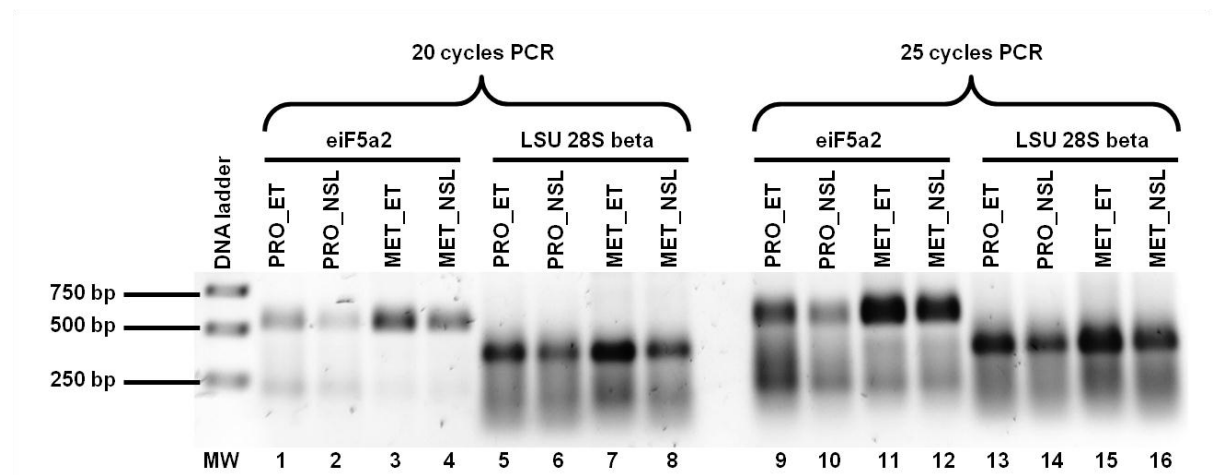


Figura 5. Poliadenilação “canônica-like” e poliadenilação truncada no 28S rRNA beta e no mRNA do eiF5a2 podem ser detectadas por oligo(dT) RT-PCR. Os RNAs poli(A)⁺ purificados e isolados das formas promastigotas de *L. (V.) braziliensis* (Lb_NSL_PRO) que estão indicados nas colunas de 1 a 16 foram usados para sintetizar a fita de DNA complementar (cDNA) com “primer” adaptador-oligo(dT). Esta reação foi seguida por PCR com um “primer” reverso que é complementar à sequência do adaptador mais um “primer” gene-específico para o gene eiF5a2 ou para o LSU 28S rRNA beta como indicado acima nas linhas horizontais. Os “amplicons” da PCR foram separados por eletroforese. As bandas fortes correspondem aos “amplicons” do mRNA do eiF5a2 com poliadenilação canônica na extremidade 3’ e aos “amplicons” do 28S rRNA com poliadenilação “canônica-like” na sua extremidade 3’. Os “smears” abaixo das bandas fortes representam a diversidade de “amplicons” truncados com tamanhos variáveis. Estes representam as moléculas de RNAs com diferentes sítios de adição de cauda rica em poli(A). A coluna do padrão molecular de DNA está representada por MW (“molecular weight”).

Analisando o alinhamento dos *reads* do transcriptoma ao gene de referência (LSU-beta) do cluster de rRNA, observamos que o tamanho da banda estava de acordo com um sítio de poliadenilação revelado por um *read* poli(A)⁺ dos dados de RNA-Seq da biblioteca transcriptômica Lb_NSL_MET, onde uma ruptura (truncagem) ocorreu antes da extremidade 3'-matura predita (Figura 6A).

A) rRNA LSU 28S-beta

```
>gi|73544646:1019295-1020819 Leishmania major strain Friedlin chromosome 27,
complete sequence
CCCAACTGCAGACCGTACTCATCACCAGCATCAGGTCACCAAGCATAGAAATGCTCTGGTACATAGAGAAAG
GTACACTCAGGGAAGTCGGCAAAATAGATGCGTAAGTTCGCAAGAAGCATTGGCTCTGAGGGCAAAGTCA
GAGAAAACCGGAAGGCCAGTGTCTCTCGCATGCAGCCGTTCTTCAGATGTTGTTTTGGCGCTCGCCGCC
TTAGCACGCTGATAGCGACTGTACTGTCGTCTGTGAAGGCTACCTTCCCAGGGCGTCGTAACCCGGCGTC
TCACTCTGGCACGGGCCGGTCTTGGCGGGTTCACGCTCGTCTTGTGCCGCGTCCCACAAATGCCAACTC
AGAAGTCTTACGGCGGGGAATCCAACGTATAAATTAACATAGGTTTGTGATGCATCCAAGTGGTGTA
TATCGCAAAGTATTTCTGCCAGTGTCTGAATGTCAACGTGACGAGATTCACCGACGCGGGGTAAC
GGCGGAGTAAGTACTGACTCTCTTAAGTAGCCAAATGCCTCGTCTTCCAATTAGAGACGCGCATGAATG
GATTAATGAGATTTCCCTCTGTCCCAGTACTATCTAGCGAAACGACAGTCAAGGGAACGACTTGAAGG
GCTAAGCGGGGAAGAAGACCCTGTTGAGTTTACTCCAGTCTGGCTCTGTGCGGCGACATCTGAGGTGT
AGTATAGTGGGAAGCGCAAGCGCAAATGAAATACCACACTCGGAACGTTGCTTCACTTATCGAATGAAG
AGACCAATGGGTTTAGCGTAGTCTTCGGGCTATGCACCGTCTAGGTTTGGGAGTAATTTTATGGCGGTAA
CCCCGTCCCCACGGGAAAGGGTGGTGTGCTTTTTTCTGCGTGCCTTTCACGGGCGTCGCGCGGGCCCCGC
TCTTTTTTATTGCGTTTCCCTTCGGGGGGCGTGACTCGTGGGTACCCAACCGTGTGTAACCTGAACCAAC
GGGATTGCCTCGTGCAGTTCGGGAGTCTTGTTCGAGACATCTGCCAGATGGGGAGTTTGGCTGGGGCGG
CATATCTGTTACACGACAACGCAAGTGTCTTAAGGCGAGCTCAGTGGGAACAGAAATCTCACGTAGAACA
CAAGCGTAAAAGCTTGCTTGATTAACGATTTCCAGTACGAATCGAGACTGCGAAAGCAAGGCCAGCGATC
CTTTGCGAAGAATAGGAAATATGAACAAATCATACCAGAGGTGTCAGAATAATTACCACAGGGATAACTG
GCTTGTGGCGGCCAAGCGTTCATAGCGACGTGCGTTTTTGTATCCTTCGATGTGCGCTCTTCCCTAACCTAG
CGCCGAGAAAGACGCTAAGGGTTGGATTGTTACCCACTGACAGGGAACGTGAGCTGGGTTTAGACCGTC
GTGAGACAGGTTGGTTTTACCCTACTTAGCTGGAAATGCGAGAAAAAGATTATCCGTGCAAAGCGGACG
TCGGTTCGTTCTTGGGCTGGCGAGCGGTGGGTAGTACTTGGTACGCTCATCACC
```

- acima mostramos a sequência completa do gene do rRNA 28S beta localizado no cromossomo 27 de *L. (L.) major* (cepa Friedlin)
- RNA-Seq de *L. (V.) braziliensis* mostrando a sequência da cauda poli(A) interna: GPST6G002H92OB
- a letra em negrito na sequência de referência: posição inicial do sítio de adição interna da cauda de poli(A)⁺ como detectado em GPST6G002H92OB
- primer direto que foi usado na RT-PCR para gerar os amplicons para validação da sequência: FW_Lb_MET_NSL_beta
- sequência do primer direto: (5') AGGGTGGTGTGCTTTTTCTG (3')
- a marcação amarela mostra o mapeamento da sequência do primer direto na sequência do gene de referência rRNA LSU 28S beta
- baseado na posição do primer direto e na sequência completa do 28S, o tamanho do amplicon esperado para a subunidade 28S-beta poliadenilada na extremidade 3' seria de 668 pb.

B) mRNA para LbrM.25.0590_eukaryotic initiation factor 5a2 putative

```
>gi|322505048:206256-206756 Leishmania braziliensis MHOM/BR/75/M2904
complete genome, chromosome 25
ATGTCTGACGAGGACCACGACTTCTCCCATCAGGGCGGCGGTGACAACGCGTCCAAGACGTACCCCTGG
CTGCTGGTGCCTTGAAGAAGGGTGGATACGTGTGCATCAACGGCCGTCCGTGTAAGGTGATTGATCTGTC
CGTGTCCAAGACCGGCAAGCATGGTTCATGCTAAGGTGAGCATTGTTGCGACCGACATCTTCACTGGCAAC
CGCCTCGAGGATCAGGCTCCATCCACGCATAACGTGGAGGTGCCGTTTGTGAAGACCTACACCTACAGCG
TGCTGGACATTCAAGCCAACGAGGATCCCTCTCTTCCAGCCACCTCTCTTTGATGGACGATGAGGGTGA
GAGCCGTGAAGACCTTGACATGCCTCCAGATCCAGCCCTGGCGACCCAGATCAAGGAACAGTTTACTCC
GGCAAGGAGGTGCTAGTCGTGGTTGTTTCTGCGATGGGCACTGAGCAGGTGCTGCAGACGAAGAATGCCG
CGGAGAAGTAG
```

- acima mostramos a sequência completa do mRNA do fator eucariótico de iniciação da tradução eiF5a2

- RNA-Seq de *L. (V.) braziliensis* mostrando a sequência da cauda poli(A) interna: GKLMG1101AS3D2

- a letra em negrito na sequência de referência: posição inicial do sítio de adição interna da cauda de poli(A)⁺ como detectado em GKLMG1101AS3D2

- primer direto que foi usado na RT-PCR para gerar os amplicons para validação da sequência: FW_Lb_MET_ET_eiF5a2

- sequência do primer direto: (5') TGCTGGTGCCTTGAAGAAG (3')

- a marcação amarela mostra o mapeamento da sequência do primer direto na sequência do gene de referência do eiF5a2

- baseado na posição do primer direto e na sequência completa do eiF5a2, o tamanho do amplicon esperado para o mRNA 3'-poli(A) canônico seria de 430 pb

Figura 6. Sítio de poliadenilação interna na subunidade beta do rRNA 28S ribossomal (read GPST6G002H92OB) e no mRNA do eiF5a2 (read GKLMG1101AS3D2) nas bibliotecas transcriptômicas de *L. (V.) braziliensis*. Poliadenilações internas nas regiões codificadoras, diferente das extremidades 3'-terminais previstas no genoma de *L. (L.) major*.

Adicionalmente, os dados do transcriptoma mostraram que um considerável número de *reads* não poliadenilados (RNA-Seq) mapearam o gene completo predito do rRNA 28S beta justamente até a posição onde a extensão poli(A) foi encontrada em um dos *reads* (Figura 7 – referente à subunidade beta do “cluster” 1 de RNAs ribossomais). Este resultado sugere um possível sítio alternativo de poliadenilação na subunidade beta madura do rRNA 28S que é menor do que a predita no genoma liberado de *L. (L.) major* (Figura 6). Por se tratar de espécies evolutivamente distantes (Momen e Cupolillo, 2000; Kerr, 2006) [sequências de *L. (V.) braziliensis* ancoradas no genoma de *L. (L.) major*] há a possibilidade de que esta subunidade, em *L. (V.) braziliensis*, seja realmente menor e o sítio de poliadenilação seja a real extremidade 3'-madura do gene. Estas hipóteses ainda necessitam de mais comprovações biológicas como, por exemplo, uma RT-PCR com “primer” reverso oligo(dT) ancorado especificamente nas regiões que flanqueiam este sítio de poliadenilação.

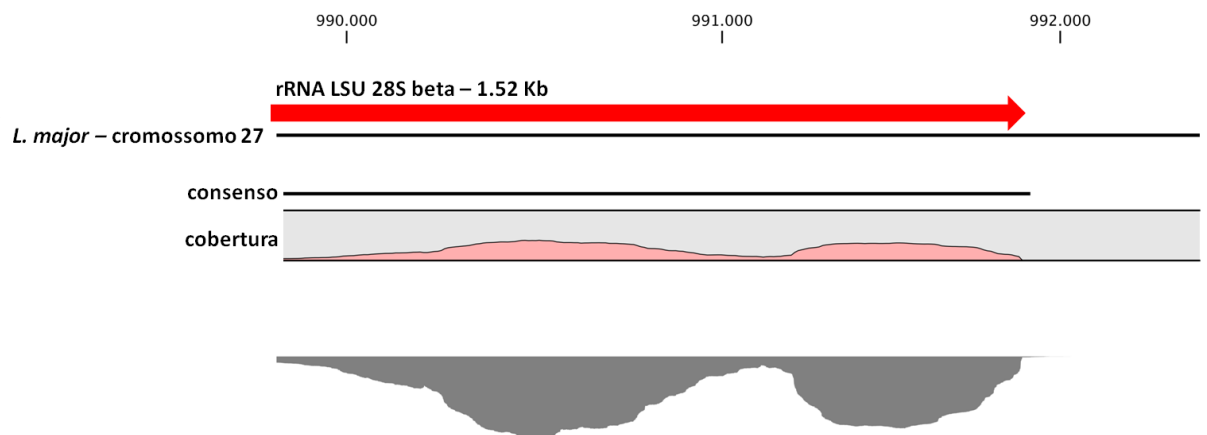


Figura 7. Representação esquemática do alinhamento de *reads* de rRNA 28S beta mostra dois conjuntos de *reads* alinhando no “template” LSU 28S-beta do primeiro cluster de rRNAs. Foram usados nesta análise todos os *reads* da biblioteca transcriptômica de Lb-PRO-NSL que alinham com a região codificante do rRNA LSU 28 beta do primeiro cluster da região de rRNA do cromossomo 27 de *L. (L.) major*, usando condições “default” do programa CLC Genomics Workbench (Version 5.0.1).

Para reforçar ainda mais a hipótese de que a poliadenilação interna não sítio-específica realmente ocorre em *Leishmania*, confirmamos por clonagem convencional e sequenciamento pelo método de Sanger que os “amplicons” dos “smears” para os rRNAs eram sequências truncadas com sítios de poliadenilação não específicos (Figura 8). Para a subunidade beta do rRNA 28S sequenciamos um clone contendo uma cauda homopolimérica interna (Figura 8A) e para o rRNA 28S alfa obtemos um clone contendo uma cauda heteropolimérica interna (Figura 8B). Adicionalmente, para o SSU 18S rRNA, nós encontramos dois clones apresentando cauda homopolimérica no mesmo ponto de ruptura (Figura 8C).

Para confirmar que este fenômeno foi também detectado em mRNAs, a presença de caudas poli(A)⁺ foi avaliada para o eiF5a2, um fator de tradução, e para o gene da proteína 60S ribossomal. Como mostrado na Figura 9, fragmentos dos genes presentes nas regiões dos “smears” que foram isolados, clonados e sequenciados também apresentaram tanto caudas homo- quanto heteropoliméricas.

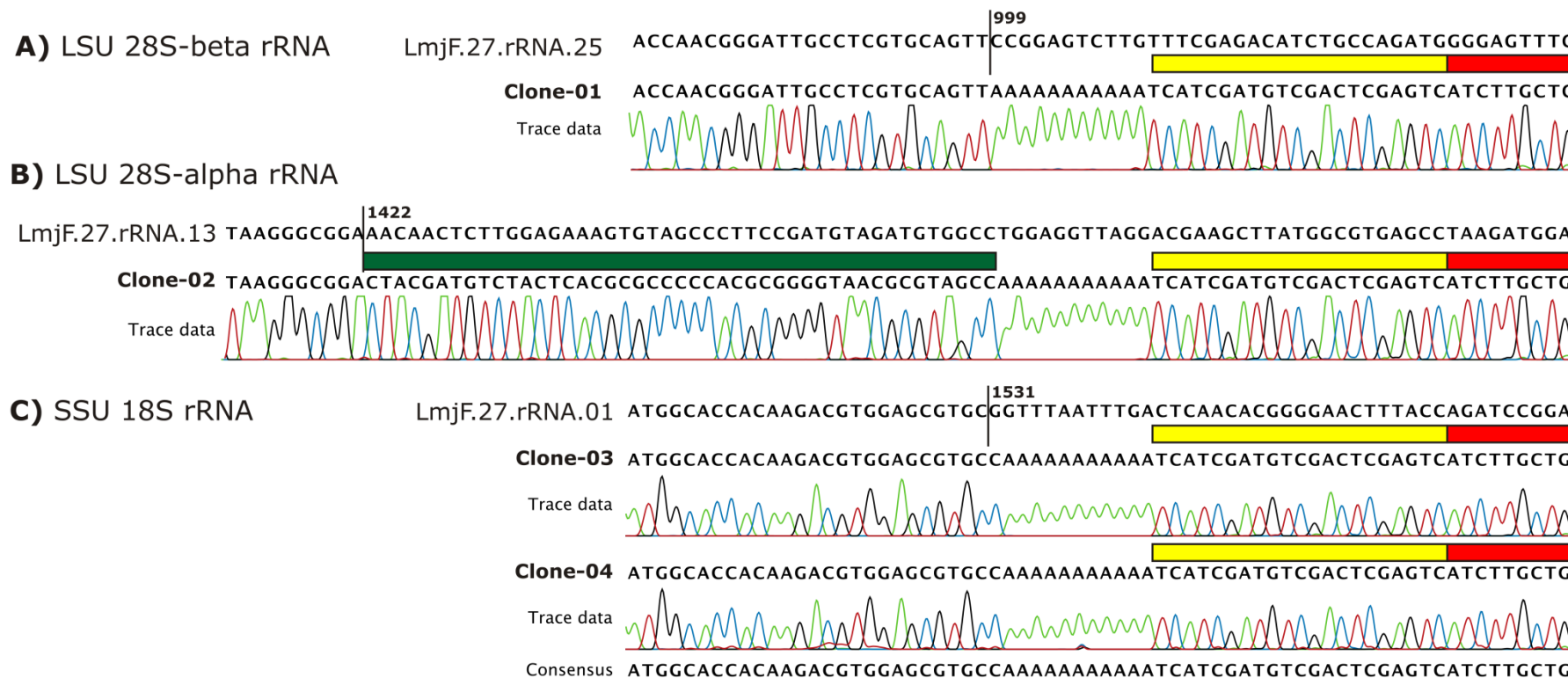


Figura 8. Moléculas de rRNAs 28S e 18S truncadas em *L.(V.) braziliensis* com caudas poli(A)⁺ homo- ou heteropoliméricas. Os cDNAs presentes nos “smears” abaixo das bandas fortes dos amplicons 3'-poli(A) canônicos dos rRNAs 28S e 18S, similares aos experimentos da Figura 5, foram purificados, clonados e sequenciados pela tecnologia de Sanger. Fragmentos da sequência de referência genômica para a subunidade beta do rRNA 28S (LmjF.27.rRNA.25), para a subunidade alfa do rRNA 28S (LmjF.27.rRNA.13) e para o SSU 18S rRNA (LmjF.27.rRNA.01) estão esquematicamente representados. Os números nas barras verticais indicam as coordenadas na sequência do ponto de ruptura onde as extensões ricas em poli(A) detectadas nos clones Sanger-sequenciados (clone 01 ao clone 04) foram adicionadas, interrompendo a sequência de rRNA. Os picos verdes no fluxograma do sequenciamento de Sanger (“Trace data”) mostram as bases de adenosina (A) detectadas nos clones. A barra verde representa a cauda heteropolimérica detectada no clone (sem alinhamento no genoma de *Leishmania*). As barras amarelas representam a sequência adaptadora do “primer” adaptador-oligo(dT) que foi usado no passo de transcrição reversa para geração dos cDNAs, e as barras vermelhas representam a sequência do vetor de clonagem.

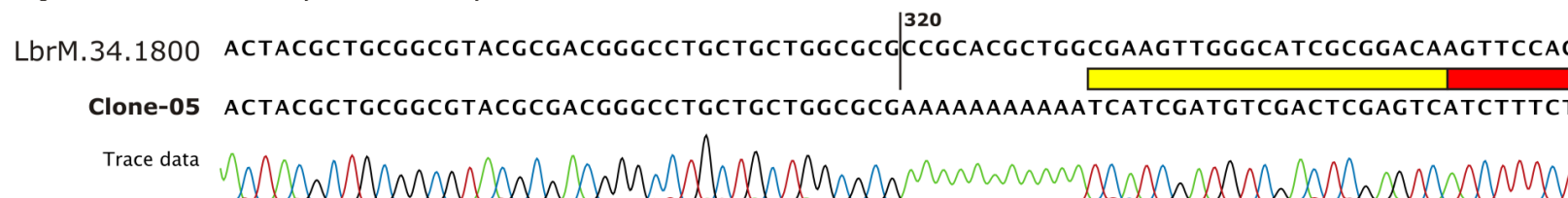
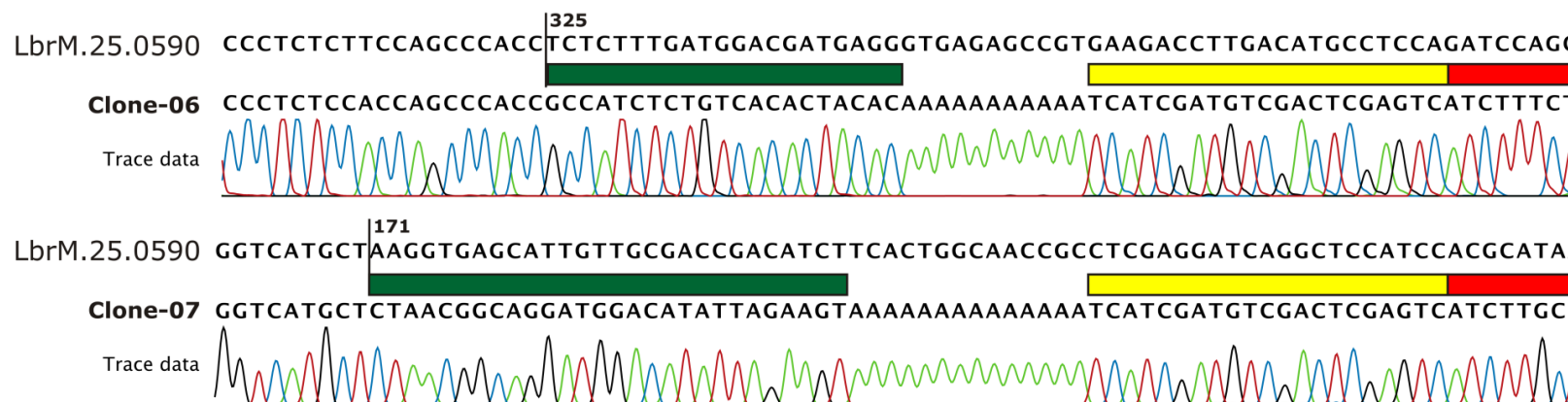
A) 60S ribosomal protein L5 putative**B) eukariotic initiation factor (eiF5a2) putative**

Figura 9. Moléculas de mRNAs (cDNA) truncadas em *L. (V.) braziliensis* com caudas poli(A)⁺ homo- ou heteropoliméricas nas mensagens da proteína 60S ribossomal L5 e do fator de tradução eiF5a2. Os cDNAs presentes nos “smears” abaixo das bandas fortes dos amplicons 3’-poli(A) canônicos, similares aos experimentos da Figura 5, foram purificados, clonados e sequenciados pela tecnologia de Sanger. Fragmentos da sequência de referência genômica para o gene da proteína 60S ribossomal L5 (LbrM.34.1800) e da sequência de referência genômica para o gene eiF5a2 (LbrM.25.0590) estão esquematicamente representados. Os números nas barras verticais indicam as coordenadas na sequência do ponto de ruptura onde as extensões ricas em poli(A) detectadas nos clones Sanger-sequenciados (clone 05 ao clone 07) foram adicionadas, interrompendo a sequência codificante. Os picos verdes no fluxograma do sequenciamento de Sanger (“Trace data”) mostram as bases de adenosina (A) detectadas nos clones. A barra verde representa a cauda heteropolimérica detectada no clone (sem alinhamento no genoma de *Leishmania*). As barras amarelas representam a sequência adaptadora do “primer” adaptador-oligo(dT) que foi usado no passo de transcrição reversa para geração dos cDNAs, e as barras vermelhas representam a sequência do vetor de clonagem.

Fragmentos Truncados Poliadenilados das Moléculas do rRNA LSU 28S beta e do mRNA do eiF5a2 Estão Presentes Tanto no Núcleo Quanto no Citoplasma de *L. (V.) braziliensis*

Como mostrado anteriormente, detectamos fragmentos de transcritos de rRNAs e mRNAs truncados poliadenilados contendo caudas homo- e heteropoliméricas. Era plausível que estes fragmentos fossem intermediários de degradação dos processos nucleares de controle de qualidade, mas uma segunda opção não poderia ser descartada – talvez a degradação auxiliada pela poliadenilação de moléculas de RNAs pudesse também ocorrer no citoplasma, coexistindo com a poliadenilação estável que caracteriza esta localização celular. Num trabalho recente, este evento de adição interna de cauda poli(A)⁺ ou rica em poli(A)⁺ foi demonstrada ocorrer no citoplasma de células humanas (Slomovic, Fremder *et al.*, 2010). Conseqüentemente, um dos nossos objetivos foi determinar se estas moléculas truncadas e poliadeniladas também estavam presentes no núcleo, no citoplasma, ou em ambos os compartimentos celulares em *Leishmania*.

Assim, as moléculas totais de RNAs foram obtidas das frações nuclear e citosólica, sendo esta última subdividida em duas categorias: (i) fração polissomal (associada aos ribossomos ativos na tradução ou polissomos) e (ii) fração pós-polissomal (solúvel ou citosólica). Para verificar a pureza das frações, a atividade enzimática da enzima citosólica glicose-6-fosfato-desidrogenase (G6PDH) e a detecção por fluorescência da presença de DNA foram mensuradas em todas as frações. Este último método foi a escolha na ausência de anticorpos específicos para detecção por imunoblot de proteínas núcleo-específicas em *Leishmania*. A Figura 10 mostra que os marcadores – DNA e proteína (G6PDH) – foram detectados exclusivamente nas frações correspondentes (frações nuclear e citosólica, respectivamente).

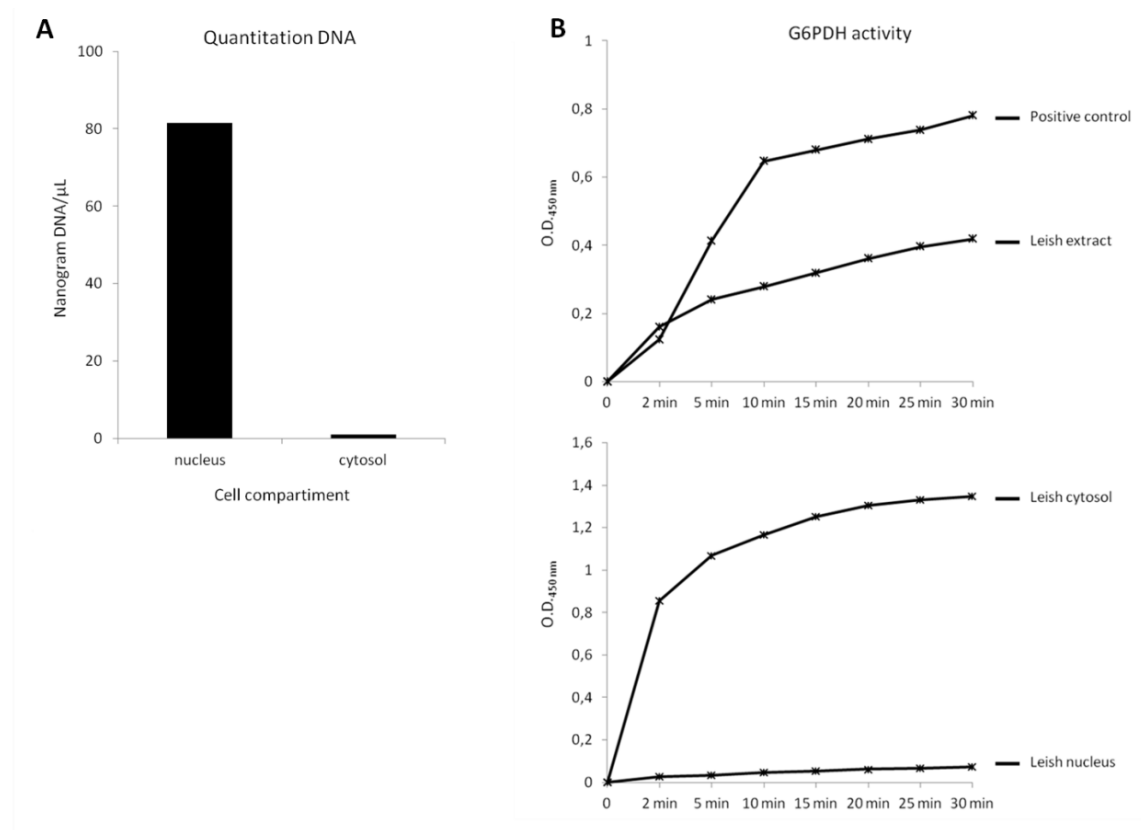


Figura 10. Quantificação do DNA e dosagem da atividade enzimática da glicose-6-fosfato-desidrogenase (G6PDH) nos compartimentos celulares de *Leishmania* após fracionamento celular. (A) Quantificação de DNA como marcador nuclear por fluorescência das frações nuclear e citosólica dos promastigotas totais do isolado NSL de *L. (V.) braziliensis* (Quant-iT™ PicoGreen dsDNA Reagent and Kits, Invitrogen/Life Technologies). (B) Quantificação da atividade da enzima G6PDH como marcadora citosólica. O painel superior representa a atividade da enzima presente no extrato total de *L. (V.) braziliensis* e a atividade do controle positivo presente no kit (G6PDH Assay Kit, BioVison Life Science). O painel inferior representa a atividade da enzima presente no extrato nuclear e no extrato citosólico de *L. (V.) braziliensis*.

Em seguida, a presença de amplicons poliadenilados do rRNA 28S beta e mRNA do eiF5a2 foram avaliados por oligo(dT) RT-PCR. Os resultados apresentados na Figura 11 mostram que para o alvo eiF5a2 observamos a banda relativa à amplificação a partir do poli(A)⁺ canônico no núcleo e no citoplasma (frações polissomal e pós-polissomal) e os rastros (“smear”) foram observados nas frações nuclear e mais fortemente na fração polissomal, mas não na fração pós-polissomal. Isso nos leva a pensar que o exossoma é muito ativo no núcleo e no citosol (como em outros organismos) e por isso não encontramos o rastro nesta última fração, já que são moléculas rapidamente degradadas. Observamos também que a maioria dos “amplicons” canônicos se encontram nas frações nuclear e polissomal, provavelmente sendo produtos de transcrição ativa e utilizados para tradução da proteína relacionada, respectivamente.

Os “amplicons” da subunidade do rRNA 28S beta apresentaram duas bandas no gel, similar aos dados do transcriptoma que mostraram um ponto de interrupção durante o alinhamento dos *reads* do LSU 28S beta baseados em dois conjuntos de *reads* agrupados justamente antes e depois da posição deste possível sítio alternativo de poliadenilação (Figura 6; Figura 7 – referente à subunidade beta do “cluster” 1 de RNAs ribossomais; Figura 11). Este resultado sugere um possível sítio alternativo de poliadenilação nesta subunidade ribossomal, diferente do predito no genoma liberado de *L. (L.) major* (Figuras 6 e 11).

Até o presente momento, esta é a primeira detecção de rRNAs e mRNAs poliadenilados internamente no citoplasma de *Leishmania*, bem como de rRNAs poliadenilados 3’ canônicos no citosol deste parasito.

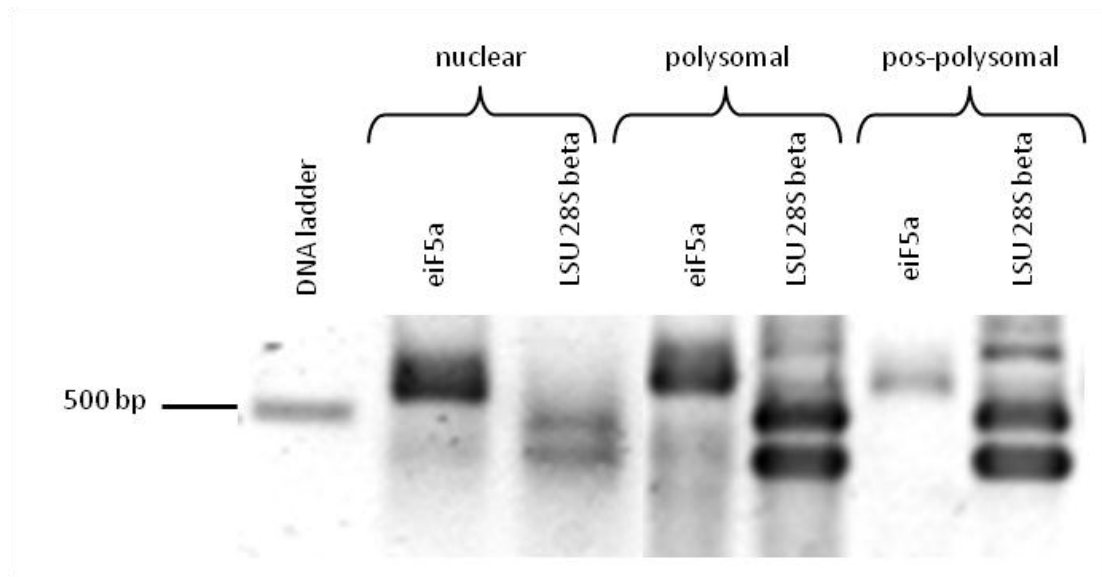


Figura 11. Eletroforese em gel de agarose 1,2% corado com brometo de etídio dos fragmentos poliadenilados amplificados por oligo(dT) RT-PCR das frações celulares dos promastigotas totais do isolado NSL de *L. (V.) braziliensis*. Após fracionamento celular, “amplicons” contendo poli(A)⁺ canônicos e internos foram amplificados após reação de transcrição reversa e PCR usando “primers” diretos específicos e “primer” reverso oligo(dT). A imagem foi digitalizada após visualização sob luz UV.

Proteínas do Complexo Exossoma Estão Presentes no Genoma de *Leishmania* e são Ativamente Transcritas

O papel do fenômeno de poliadenilação interna em *Leishmania* é desconhecido porque esta foi a primeira observação em dados biológicos. Para reforçar a hipótese de que moléculas de RNAs truncadas poderiam ser direcionadas para as vias de degradação de RNAs, como observado em outros organismos, resolvemos analisar se *Leishmania* possui em seu genoma subunidades similares àquelas do exossoma de eucariotos superiores e, se caso tivessem, se eram ativamente transcritas. Para isto, fizemos uma busca na literatura para definir quais seriam as subunidades de exossoma a serem procuradas.

Assim, analisamos as bibliotecas transcriptômicas sequenciadas dos isolados ET e NSL de *L. (V.) braziliensis* e as proteínas preditas no Tritryp e encontramos que 90% dos ortólogos do exossoma de mamíferos e leveduras estão presentes em *L. (V.) braziliensis* e são expressos nas bibliotecas de transcriptoma (Tabela 6). Portanto, consideramos a possibilidade de que o fenômeno de poliadenilação transiente de RNAs truncados em *Leishmania* seja amplamente disperso nos parasitos da ordem Kinetoplastida, uma vez que esta maquinaria também já foi descrita no *T. brucei* (Estevez, Kempf *et al.*, 2001) e pode estar relacionada com processos de controle de qualidade das moléculas de RNAs expressas num determinado momento nestes parasitos.

Tabela 6. Genes ortólogos do complexo do exossoma identificados nos dados de transcriptoma de *L. (V.) braziliensis* (isolados ET e NSL) e no TriTrypDB.

UniProt ID	Nome da proteína	TritypDB ID	Proteína em <i>L.(V.) braziliensis</i> (resultado do BlastX)	E-value	Reads mapeados	Cobertura (%)
Q9NQT5	Exosome complex component RRP40 ^a	LbrM.04.0170	exosome complex exonuclease RRP40, putative	1.1e ⁻⁰⁸	41	100
Q8IZH2	5'-3' exoribonuclease 1 (Xrn1) ^a	LbrM.06.0240	5'-3' exonuclease, putative	1.6e ⁻¹²⁶	344	100
P53632	Poly(A) RNA polymerase protein 2 (Trf4) ^b	LbrM.07.0770	topoisomerase-related function protein-like protein	4.6e ⁻²⁴	178	100
Q13868	Exosome complex component RRP4 ^a	LbrM.14.0010	exosome complex exonuclease, putative, ribosomal RNA processing protein 4, putative	3.4e ⁻³⁷	31	100
Q01780	Exosome component 10 (PM/ScI-100) ^a	LbrM.20.2660	exosome subunit rrp6p homologue, putative	6.2e ⁻⁷³	73	100
Q96B26	Exosome complex component RRP43 ^a	LbrM.22.1490	exosome complex exonuclease RRP45 homolog, putative	7.2e ⁻²²	74	100
Q06265	Exosome complex component RRP45 (PM/ScI-75) ^a	LbrM.22.1490	exosome complex exonuclease RRP45 homolog, putative	6.2e ⁻²⁶	74	100
Q15024	Exosome complex component RRP42 ^a	LbrM.22.1490	exosome complex exonuclease RRP45 homolog, putative	1.6e ⁻⁰⁸	74	100
Q9Y3B2	Exosome complex component CSL4 ^a	LbrM.23.1670	Csl4p homologue, putative	0.00011	13	100
Q9Y2L1	Exosome complex exonuclease RRP44 (Dis3) ^a	LbrM.28.0340	rrp44p homologue, putative	3.0e ⁻¹⁵⁷	252	100
Q8TF46	DIS3-like exonuclease 1 (Dis3L1) ^a	LbrM.28.0340	rrp44p homologue, putative	7.4e ⁻¹¹⁹	252	100
P40507	Protein AIR1 ^b	LbrM.35.1770	universal minicircle sequence binding protein	1.3e ⁻¹⁰	682	100
Q12476	Protein AIR2 ^b	LbrM.35.1770	universal minicircle sequence binding protein	3.5e ⁻¹¹	682	100
P47047	ATP-dependent RNA helicase DOB1 (Mtr4) ^b	LbrM.35.3220	ATP-dependent RNA helicase, putative	1.9e ⁻²⁰³	120	100
Q9NPD3	Exosome complex component RRP41 ^a	LbrM.35.3240	exosome complex exonuclease rrp41, putative	4.1e ⁻³²	35	100
Q5RKV6	Exosome complex component MTR3 ^a	LbrM.35.5140	exosome-associated protein 4, putative	4.1e ⁻⁰⁵	23	100
Q9NQT4	Exosome complex component RRP46 ^a	LbrM.35.5140	exosome-associated protein 4, putative	0.00018	23	100

^aSequências de *Homo sapiens* usadas como “template” – complexo do exossoma. ^bSequências de *S. cerevisiae* – complexo TRAMP. Os reads mapeados são a somatória dos reads dos isolados ET e NSL de *L. (V.) braziliensis*.

DISCUSSÃO

O desenvolvimento de novas metodologias de sequenciamento em larga escala de alto rendimento forneceu novos métodos para o mapeamento e quantificação do transcriptoma (Lao, Tang *et al.*, 2009; Reis-Filho, 2009). Um dos maiores benefícios do RNA-Seq é sua alta capacidade de cobertura (Wang, Gerstein *et al.*, 2009).

Neste trabalho, inicialmente queríamos estudar as diferenças na expressão gênica global em larga escala entre dois isolados com características biológicas marcadamente distintas para compreender melhor os mecanismos de infectividade e virulência em *L. (V.) braziliensis*. Para isto, o primeiro passo foi a construção das bibliotecas transcriptômicas a partir dos mRNAs poli(A)⁺, mas no decorrer do estudo encontramos um persistente alto conteúdo de moléculas de rRNAs nas amostras de mRNAs após o passo de purificação por afinidade com “beads” magnéticos conjugados com oligo(dT) (Figura 2). Então, suspeitamos da existência de pequenas extensões internas ricas em poli(A) nestes rRNAs supostamente “contaminantes” em relação aos mRNAs e tentamos nos livrar dos mesmos pelo aumento da temperatura durante o passo de purificação das amostras poli(A)⁺, mas sem qualquer sucesso (Figuras 1, letras C e D; Tabela 2). Estes resultados sugeriram que os rRNAs poderiam ter caudas poli(A)⁺ canônicas, similar ao que é amplamente conhecido para mRNAs.

Para testar esta possibilidade, sequenciamos e analisamos os transcritos de rRNAs presentes nas bibliotecas transcriptômicas. Considerando as cinco bibliotecas sequenciadas [quatro de *L. (V.) braziliensis* e uma de *L. (L.) amazonensis*], nosso mapeamento revelou 1.131.543 *reads* homólogos alinhados no “template” do “cluster” gênico de rRNAs no genoma (rDNA) de *L. (L.) major*. Este número corresponde a quase 25% do total de sequências obtidas. De fato, nosso agrupamento mapeou todos os 52 genes de rRNAs, e eles foram mapeados com 100% de cobertura em relação ao “cluster” de *L. (L.) major*.

Encontramos *reads* de rRNAs poliadenilados na extremidade 3'-matura das subunidades, na extremidade 3'-não processada e em regiões internas das subunidades maduras (Figuras 2, 3 e 4; Tabela 4). Estes dados sugerem fortemente que estas subunidades do rRNA contém uma cauda poli(A)⁺ na extremidade 3'-terminal similar à cauda poli(A)⁺ canônica encontrada nos mRNAs. Adicionalmente,

estes dados mostram que a poliadenilação do rRNA poderia ocorrer co-transcricionalmente (rRNA pré-processado) e, ou, pós-transcricionalmente (rRNA maduro). Entretanto, outras abordagens são necessárias para investigar estas hipóteses.

Apesar da relativa distância evolutiva entre os dois subgêneros em *Leishmania* (Momen e Cupolillo, 2000; Kerr, 2006), a poliadenilação do rRNA parece ser um evento comum e amplamente disperso entre os dois subgêneros uma vez que este evento foi previamente observado em duas diferentes espécies, *L. (V.) braziliensis* e *L. (L.) infantum* (Decuypere, Vandesompele *et al.*, 2005), assim como por nós também observado em *L. (V.) braziliensis* e *L. (L.) amazonensis* (Figura 3; Tabela 4). Porém, a literatura anterior evidenciou somente a poliadenilação da última subunidade (épsilon) do transcrito policistrônico de rRNAs e nossos dados mostram que todas as subunidades parecem ser poliadeniladas na porção 3', assim como ocorre com os mRNAs.

Considerando que originalmente nossos dados de eletroforese capilar evidenciam que todos os rRNAs devem ser poliadenilados, um alto conteúdo de caudas poli(A)⁺ ou ricas em poli(A)⁺ na extremidade 3'-terminal dos *reads* de transcriptoma era esperado. Porém, o que observamos foi um baixo número de *reads* com uma cauda poli(A)⁺ evidente (Tabelas 4). Estes resultados poderiam ser explicados pela abordagem experimental usada na construção das bibliotecas transcriptômicas. Durante a fragmentação dos mRNAs em pontos aleatórios e síntese dos cDNAs com “primers” randômicos, as extremidades 3'-terminais contendo as caudas poli(A)⁺ originais podem ter sido perdidas, como observado pela baixa abundância de *reads* poliadenilados. O mesmo efeito poderia estar ocorrendo com as poliadenilações truncadas porque elas são internas em relação ao tamanho total da região codificadora e foram adicionadas interrompendo a informação dos rRNAs e mRNAs. Entretanto, em relação ao transcrito em si elas estão na extremidade 3'-terminal destes.

Decuypere e colaboradores (2005) demonstraram que os parasitos do gênero *Leishmania* são capazes de poliadenilar o precursor do rRNA em níveis variáveis em diferentes cepas e em diferentes estágios de vida de uma mesma cepa, sendo esta poliadenilação presente na porção 3' da última subunidade (épsilon) do transcrito

policistrônico de rRNAs, indicando que este evento ocorre durante o processo de maturação dos rRNAs (Decuypere, Vandesompele *et al.*, 2005). Inicialmente, os autores buscavam diferenças na expressão gênica que pudessem explicar as diferenças biológicas entre cepas naturalmente resistentes e sensíveis ao antimônio pentavalente através da comparação de cDNAs produzidos por oligo(dT) RT-PCR usando supressão subtrativa. De modo inesperado, um número substancial de genes candidatos diferencialmente expressos foram representados por fragmentos do rRNA LSU 28S épsilon, indicando que os rRNAs em *Leishmania* eram poliadenilados. O passo inicial foi verificar se o rRNA LSU 28S continha uma cauda 3' poli(A) e qual era a extensão deste fenômeno. Os resultados claramente demonstraram que uma parte significativa do precursor da subunidade épsilon do LSU 28S rRNA sofre extensiva poliadenilação pós-transcricional.

De modo interessante, detectamos nos perfis de eletroforese dos RNAs poli(A)⁺ purificados de *L. (V.) braziliensis* (cepa NSL) três picos mais intensos que provavelmente correspondem aos rRNAs SSU 18S, LSU 28S alfa e LSU 28S beta (Figura 2), porque os tamanhos calculados destas subunidades de rRNA estão de acordo com as previsões obtidas de dados genômicos para *L. (L.) major* e outros tripanosomatídeos (El-Sayed, Myler, Blandin *et al.*, 2005; Ivens, Peacock *et al.*, 2005). Resultados similares foram obtidos para as diferentes formas evolutivas das demais cepas analisadas também (Figura suplementar S1). Isto sugere que todas as subunidades dos rRNAs poderiam ser poliadeniladas na extremidade 3'-terminal, e não exclusivamente o precursor da subunidade épsilon, como descrito previamente (Decuypere, Vandesompele *et al.*, 2005).

Analisando as extremidades 3' dos *reads* poliadenilados não abundantes derivados das subunidades do rRNA nas bibliotecas transcriptômicas, observamos que muitos deles continham caudas homo- ou heteropoliméricas internamente localizadas (Figuras 3 e 4; Tabela 4). Uma observação similar da coexistência de caudas homopoliméricas ou heteropoliméricas ricas em poli(A) em moléculas truncadas de rRNAs e mRNAs foi recentemente descrita para fragmentos de RNAs no núcleo e no citoplasma de células humanas (Slomovic, Fremder *et al.*, 2010). De maneira similar, identificamos “amplicons” truncados transcritos contendo extensões poliadeniladas no núcleo e no citoplasma em *Leishmania*. A acumulação não

resultou somente em uma única banda de forte intensidade [representando os sítios canônicos 3'-poli(A)⁺], mas mostrou também um rastro de “amplicons” (“smear”), logo abaixo da respectiva banda forte de cada gene utilizado como modelo para validação do fenômeno (Figuras 5 e 11). Até o presente momento, esta é a primeira detecção de rRNAs e mRNAs poliadenilados internamente no citoplasma de *Leishmania*. É bem estabelecido e amplamente conhecido que os mRNAs são estavelmente poliadenilados na extremidade 3'-terminal. Contudo, a poliadenilação transiente interna de RNAs parece estar diretamente relacionada com mecanismos de controle de qualidade de RNAs recentemente descritos em um número ainda limitado de organismos eucarióticos (Fleischmann, Liu *et al.*, 2004; Kuai, Fang *et al.*, 2004; Kao e Read, 2005; Slomovic, Laufer *et al.*, 2006). Nossos dados sugerem que estes eventos também ocorrem em *Leishmania*.

A confirmação da poliadenilação nas extremidades 3'-matura dos rRNAs e da adição interna de caudas poli(A)⁺ e ricas em poli(A)⁺ em rRNAs e mRNAs foram realizadas por RT-PCR e sequenciamento de Sanger (Figuras 5, 8 e 9). A presença de caudas 3'-poli(A)⁺ canônicas gerou um perfil de bandas específicas e as poliadenilações internas produziram produtos de amplificação dispersos (“smears”) pelo gel de agarose, logo abaixo das respectivas bandas canônicas. Surpreendentemente, a banda amplificada relacionada com a poliadenilação canônica da subunidade beta do LSU 28S rRNA, usado nas validações, apresentou um tamanho menor do que o esperado ao que foi calculado usando-se o rRNA relativo de *L. (L.) major* e o consenso do rRNA respectivo de *L. (V.) braziliensis* gerado pela análise dos dados de transcriptoma deste trabalho. Este resultado sugere a presença de um sítio alternativo de poliadenilação anterior ao fim da região codificante para este rRNA. Está bem claro na literatura que os mRNAs tem uma variação na extensão da sua 3'-UTR de aproximadamente cinquenta a algumas centenas de nucleotídeos com um sítio de poliadenilação clássico (3'-AAUAAA) ou algumas variações desta sequência (Proudfoot, 2011). Nossos dados sugerem que um evento similar possa ocorrer na transcrição da subunidade épsilon do LSU 28S rRNA, apesar das características dispersas dos sítios de poliadenilação nestas moléculas (Lacava, Houseley *et al.*, 2005).

Observamos longas caudas homo- ou heteropoliméricas internas em todas as bibliotecas transcriptômicas analisadas, mas somente caudas curtas foram determinadas pelo sequenciamento de Sanger, tanto em moléculas de rRNAs quanto de mRNAs (Figuras 8 e 9). Em contraste aos nossos dados de Sanger, longas caudas poli(A)⁺ foram identificadas nos intermediários de degradação em células humanas (Slomovic, Laufer *et al.*, 2006; Slomovic, Fremder *et al.*, 2010). Há algumas possíveis explicações para estas diferenças. A primeira é que o tempo de vida transiente dos RNAs truncados poliadenilados durante o isolamento e clonagem dos produtos de amplificação poderia levar à degradação parcial das caudas poli(A)⁺, similar às descrições anteriores em outros modelos (Kuai, Fang *et al.*, 2004; Lacava, Houseley *et al.*, 2005; Slomovic, Fremder *et al.*, 2010). Como não utilizamos a circularização dos cDNAs antes de realizar a etapa de PCR das RT-PCRs, durante o processamento das amostras, naturalmente as caudas longas podem ter sido degradadas. Assim, uma alternativa para evitar esta degradação e obter caudas mais longas e estabilizadas seria usar esta etapa de circularização que é sugerida como parte do protocolo para análise deste tipo de processamento de RNAs (Slomovic, Portnoy *et al.*, 2008).

Outra explicação estaria relacionada com uma possível baixa processividade da enzima poli(A) polimerase envolvida com este fenômeno em *Leishmania*. É sabido que durante o processamento co-transcricional do mRNA, a Pap1p é responsável pela adição da cauda 3'-poli(A)⁺ seguinte ao processo de clivagem (Hirose e Manley, 1998; Zhao, Hyman *et al.*, 1999; Proudfoot e O'sullivan, 2002; Proudfoot, 2011). É possível que parte das diferenças funcionais pudesse surgir da velocidade e processividade das reações de poliadenilação, mas este campo ainda é totalmente desconhecido em *Leishmania*. Em eucariotos superiores, a Pap1p é altamente processiva e rapidamente adiciona aproximadamente 250 As (mRNAs de humanos) e 90 As aos mRNAs de leveduras. Esta atividade processiva presumivelmente previne o acesso da extremidade 3'-terminal dos substratos de RNAs ao complexo do exossoma até que a polimerização esteja completa e o transcrito seja estabilizado pela ligação das proteínas com domínio de ligação à cauda poli(A)⁺ (Pab1p).

Em leveduras, a enzima central (Trf4p) do complexo auxiliar de poliadenilação nuclear, chamado complexo TRAMP, é responsável pela poliadenilação de moléculas de RNAs truncados e é específica para nucleotídeos de adenosina quando analisada *in vitro* (Haracska, Johnson *et al.*, 2005; Lacava, Houseley *et al.*, 2005; Vanacova, Wolf *et al.*, 2005; Wyers, Rougemaille *et al.*, 2005; Egecioglu, Henras *et al.*, 2006). De modo interessante, somente caudas homopoliméricas foram encontradas em *S. cerevisiae*, indicando que o componente responsável pela adição de caudas heteropoliméricas não está presente neste organismo (Slomovic, Fremder *et al.*, 2010). Algumas poli(A) polimerases adicionam As ou Us, entretanto, muitas das sequências heteropoliméricas encontradas em nossos resultados foram ricas em A/G e poderiam ser adicionadas por outra enzima, como descrito em outros organismos (Mullen e Marzluff, 2008; Heo, Joo *et al.*, 2009). Além do mais, é improvável que estas moléculas de rRNAs truncadas poliadeniladas sejam processadas por estas enzimas canônicas (Pap1p) por causa das características dispersas dos sítios de poliadenilação ao longo das moléculas de rRNAs, como já dito anteriormente (Lacava, Houseley *et al.*, 2005).

Diante de toda esta discussão, vemos que a poliadenilação interna de moléculas de rRNAs é uma evidência da presença de uma via de degradação poli(A)-estimulada que serve como um sistema de controle de qualidade na rápida remoção de moléculas que são mal processadas ou incorretamente dobradas, e que não estão diretamente associadas com os ribossomos. Outra possibilidade é que os fragmentos transcritos poliadenilados poderiam ser intermediários poliadenilados da degradação de RNAs originados de rRNAs maduros e funcionais localizados no citoplasma, mas a poli(A) polimerase relacionada não pôde ser identificada (Slomovic, Laufer *et al.*, 2006; Slomovic, Fremder *et al.*, 2010). Slomovic e colegas (2010) tentaram, sem sucesso, identificar as poli(A) polimerases responsáveis pela adição das caudas homo- ou heteropoliméricas internas em células humanas. Somente conseguiram afirmar que as poli(A) polimerases canônicas (Pap1p) conhecidas não são as responsáveis por estes eventos, tornando assim, um campo aberto para pesquisas futuras (Slomovic, Fremder *et al.*, 2010). Essas questões demonstram a importância e necessidade de mais experimentações para clarear o entendimento dos elementos envolvidos no complexo de poliadenilação e nos mecanismos de degradação.

Nossos dados também sugerem que a poliadenilação interna de rRNAs e mRNAs pode ocorrer na fração polissomal de ribossomos ativos (Figura 11), sugerindo talvez a possibilidade de que rRNAs e mRNAs utilizados na tradução possam ser controladamente levados à degradação, provavelmente após seu uso. É conhecido que as caudas poli(A)⁺ realizam um papel importante na modulação da estabilidade dos RNAs em muitos sistemas. Em vários organismos, uma poliadenilação transiente em moléculas truncadas ou mal dobradas atua como parte de um sistema de decaimento de RNA (Kushner, 2004). No citoplasma de eucariotos, as caudas 3'-poli(A) canônicas e as proteínas ligantes de poli(A) aumentam a estabilidade das moléculas de mRNAs (Wilusz, Gao *et al.*, 2001; Mangus, Evans *et al.*, 2003). Em contraste, em bactérias, a cauda poli(A) age como elemento desestabilizador (Deutscher, 2006).

Nas mitocôndrias, os efeitos da poliadenilação diferem dramaticamente de organismo para organismo. Por exemplo, na mitocôndria humana, foi observada a coexistência de transcritos truncados poliadenilados não abundantes juntos com os caracterizados com cauda poli(A) estável (Slomovic, Laufer *et al.*, 2005). Em *T. brucei*, a poliadenilação estimula a degradação de moléculas de RNAs não editados presentes na mitocôndria. Neste estudo, a sequência primária de quatro versões diferentemente editadas do mesmo substrato de mRNA foi usada – (i) o mRNA não editado, (ii) parcialmente editado em seis sítios na extremidade 3'-terminal, (iii) parcialmente editado em 45 sítios e (iv) totalmente editado (77 sítios de edição). Para determinar se as sequências editadas atuavam como elementos regulatórios *cis*-atuantes no decaimento da molécula de RNA, à parte da cauda 3'-poli(A), foi comparada a taxa de degradação dos RNAs não editados e dos RNAs completamente editados, ambos não poliadenilados. Estes últimos degradaram significativamente mais rápidos do que sua contrapartida não editada (Kao e Read, 2005). Também foi determinado o grau de edição necessário para facilitar a rápida degradação dos transcritos. Os RNAs parcialmente editados com 45 sítios de edição apresentaram uma taxa de degradação mais rápida do que os RNAs completamente editados, sugerindo que a completa edição não é requerida para facilitar o rápido decaimento dos substratos de RNAs não poliadenilados. Surpreendentemente, os RNAs parcialmente editados com apenas seis sítios de edição (adição de 6 uridinas e deleção de outras três), no qual se modificou levemente a extremidade 3'-terminal da

molécula, desestabilizou fortemente o transcrito de maneira similar às outras edições. Estes resultados demonstraram que as sequências editadas *cis*-atuantes, compreendendo somente seis sítios de edição, são suficientes para desestabilizar os transcritos mitocondriais em *T. brucei* (Kao e Read, 2005).

Num estudo anterior do mesmo grupo de pesquisadores foi demonstrado que uma cauda poli(A)₂₀ serve como elemento desestabilizante na extremidade 3' de mRNAs não editados (Ryan, Militello *et al.*, 2003). Contrastando fortemente com estes resultados, a cauda poli(A)₂₀ impediu o decaimento desta mesma molécula quando parcial ou completamente editada (Kao e Read, 2005). Estes experimentos demonstraram que a cauda 3'-poli(A) tem efeitos opostos na estabilidade de um mesmo mRNA quando editado ou não. A demonstração de que a cauda poli(A)₂₀ estabiliza os transcritos editados sugere que estes RNAs editados poliadenilados possam ser protegidos por proteínas ou um complexo de proteínas que reconhece a cauda poli(A) conjuntamente com as sequências editadas *cis*-atuantes (Kao e Read, 2005).

O tamanho da extensão da cauda poli(A) também é importante para determinar a estabilidade dos RNAs editados. De um modo geral, os RNAs contendo uma cauda poli(A)₂₀₀ degradam mais lentamente do que as moléculas completamente editadas não poliadeniladas, mas ligeiramente mais rápido do que as moléculas com cauda poli(A)₂₀. Assim, as longas caudas poli(A) aparentemente têm um moderado efeito sobre a estabilidade dos RNAs mitocondriais editados (Kao e Read, 2005). Outro importante fator determinante da estabilidade das moléculas de RNAs mitocondriais é a composição da cauda poli(A). Os mRNAs mitocondriais de *T. brucei* contém resíduos de uridina interespaçando os resíduos de adenosina na cauda poli(A) (Decker e Sollner-Webb, 1990; Souza, Myler *et al.*, 1992). Ficou demonstrado que as caudas poli(A)₂₀ contendo resíduos de uridinas intercalados também impedem o decaimento dos mRNAs mitocondriais editados, similar às extensões homopoliméricas de adenosina. Entretanto, caudas poli(A) muito curtas (5 resíduos de adenosina) ou exclusivamente compostas por outro nucleotídeo (20 resíduos de uridina) são incapazes de mediar o lento decaimento dos RNAs mitocondriais editados. Em conjunto, todos estes resultados evidenciaram que tanto o tamanho quanto a composição de nucleotídeos são importantes para a estabilização

mediada pela cauda poli(A) dos RNAs mitocondriais editados em *T. brucei* (Kao e Read, 2005).

Adicionalmente, um mecanismo de controle de qualidade que realiza uma poliadenilação transiente (complexo TRAMP) e auxilia na degradação de RNAs alvos pelo complexo do exossoma nuclear foi bem descrito em *S. cerevisiae* (Lacava, Houseley *et al.*, 2005; Vanacova, Wolf *et al.*, 2005; Wyers, Rougemaille *et al.*, 2005). Recentemente, também foi demonstrado que o silenciamento (“Knockdown”) através de interferência por RNA (RNAi) de alguns componentes do complexo do exossoma permitiu a acumulação dos RNAs poliadenilados truncados no citoplasma de células humanas, sugerindo o envolvimento do exossoma neste processo (Slomovic, Fremder *et al.*, 2010).

Identificamos em nossos resultados sequências de genes ortólogos ao complexo do exossoma e ao complexo auxiliar (TRAMP) de humanos e leveduras, suportando a hipótese de que este fenômeno também está presente em *Leishmania* (Tabela 6). A identificação nos nossos resultados de transcritos ortólogos nas bibliotecas de transcriptoma, com possível homologia funcional aos componentes do exossoma de leveduras e humanos, e com provável participação num complexo de degradação, é reforçada pela comprovação experimental da presença do exossoma em *T. brucei*, organismo taxonomicamente próximo de *Leishmania* (Estevez, Kempf *et al.*, 2001). Os dados deste trabalho evidenciaram a existência de um complexo de degradação mais simplificado, que possui componentes com atividade processiva 3'→5' ribonuclease, essenciais para o crescimento celular e comprometidos na maturação do rRNA 5.8S. Os autores identificaram oito genes codificantes para proteínas similares aos componentes do exossoma em leveduras (Estevez, Kempf *et al.*, 2001). Apesar disso, o número de componentes do exossoma em *T. brucei* pode ser um pouco maior uma vez que, na época da comprovação experimental, o genoma deste patógeno ainda não estava completamente finalizado.

A simplicidade do exossoma neste organismo sugere que este complexo tornou-se maior e mais elaborado durante a evolução eucariótica. Apesar das diferenças na composição das subunidades, o modo de ação geral do exossoma parece ser bastante similar entre *T. brucei* e *S. cerevisiae*. Além do mais, alguns dos componentes do exossoma identificados atuam como proteínas separadas, indicando

que exonucleases individuais podem ter começado a se associar em um grande complexo antes ou durante o aparecimento da linhagem eucariótica. Com o aumento da complexidade dos organismos, o processamento de RNAs e sua regulação tornaram-se mais complicados, e mais subunidades foram adicionadas (Estevez, Kempf *et al.*, 2001). Assim, o exossoma simplificado neste organismo fornece uma visão pontual dos processos evolutivos desses eventos biológicos.

De um modo geral, o papel da poliadenilação em rRNAs e mRNAs truncados, como observado aqui neste presente trabalho realizado em *Leishmania*, abre novos campos de pesquisa para determinar o real impacto deste fenômeno sobre o mecanismo da regulação gênica nestes parasitos e nas espécies relacionadas. A capacidade muito limitada de controlar a transcrição de genes individuais em tripanosomatídeos, com uso de poucos fatores de transcrição diferentes, limita muito a regulação da expressão gênica nestes organismos em nível de transcrição (Clayton, 2002; Clayton e Shapira, 2007; Haile e Papadopoulou, 2007). Isto levou vários autores a postularem que tripanosomatídeos transcrevem seus genomas praticamente de forma constitutiva. Desta forma, para sua adaptação e sobrevivência em diferentes ambientes, postula-se que a regulação da expressão gênica nestes organismos seja quase exclusivamente ou muito mais importante e finamente regulada em nível pós-transcricional (Clayton e Shapira, 2007). Assim, acredita-se que a regulação da expressão gênica nestes organismos seja feita principalmente através da regulação do tempo de vida dos mRNAs, bem como pela estabilidade e integridade das demais moléculas estruturais dos RNAs não codificadores e através de regulação em nível de tradução.

A poliadenilação em moléculas de rRNAs também foi demonstrada em *Candida albicans*, um dos fungos patogênicos mais comuns que afetam humanos, principalmente os indivíduos imunocomprometidos (Edmond, Wallace *et al.*, 1999; Fleischmann e Liu, 2001; Fleischmann, Liu *et al.*, 2004). A mudança para a forma germinativa neste organismo é estimulada pelo soro sanguíneo dos pacientes infectados. Num estudo inicial dos genes envolvidos na transição celular para a forma de hifa, surpreendentemente as análises de cDNAs revelaram que tanto a forma celular quanto a germinativa deste organismo poliadenila alguns de seus transcritos de rRNA 25S (Fleischmann e Liu, 2001). Neste mesmo trabalho ficou

demonstrado que há um rápido e transiente aumento do processo de poliadenilação após a estimulação deste fungo pelo soro humano. Estes dados indicam que a poliadenilação do rRNA 25S é parte do processo biológico do desenvolvimento celular em *C. albicans* e possivelmente regula a transição entre formas evolutivas neste organismo (Fleischmann e Liu, 2001).

Num trabalho mais recente, os mesmos autores mostraram que este processo também envolve a subunidade menor 18S do rRNA neste organismo. Do mesmo modo como no rRNA 25S, os rRNAs 18S poliadenilados se tornaram mais abundantes logo após o contato com o soro e aparecimento das hifas (Fleischmann, Liu *et al.*, 2004). Contrariamente ao que ocorre para a subunidade maior 25S do rRNA, foi mostrado que os sítios de poliadenilação no rRNA 18S estão localizados próximos da extremidade 3'-terminal e são mais variáveis, mas nenhuma poliadenilação foi encontrada exclusivamente nos sítios de maturação descritos para esta molécula (Fleischmann, Liu *et al.*, 2004). Talvez a extremidade 3' da subunidade 18S realize um importante papel no reconhecimento do sítio inicial do mRNA, sendo vigorosamente protegida de modificações. Isto sugere que a poliadenilação no rRNA 18S possa realizar uma função adicional no ribossomo.

A mudança para a forma germinativa (hifa ou pseudohifa) ocorre sempre após uma variedade de estímulos externos, submetendo as células a fortes estresses fisiológicos. Nestas condições, os rRNAs em *C. albicans* poderiam ser poliadenilados para rapidamente regular a expressão gênica quando necessária através do seu ciclo de vida (Fleischmann, Liu *et al.*, 2004). Em *Leishmania*, a abundância e a qualidade destas moléculas de RNAs podem ser finamente controladas e correlacionadas com a modulação das condições fisiológicas e bioquímicas por todo o seu ciclo de vida. De maneira similar, nossos resultados sugerem que os fragmentos truncados poliadenilados dos transcritos possam servir como alvos para mais um tipo de regulação da expressão gênica.

Há evidências em outros organismos, já discutidas anteriormente, que nos levam a supor que a adição de caudas poli(A) facilite o decaimento destas moléculas truncadas, indicando que a poliadenilação possa agir como um mecanismo pelo qual as células rapidamente desestabilizam, quando necessário, genes individuais ou um conjunto de genes com funções relacionadas. Porém, também vimos evidências de

que o fenômeno ocorra co-traducionalmente, sugerindo que mRNAs e rRNAs utilizados na tradução também possam ser truncados por poliadenilação não sítio-específica. Ainda são necessárias mais pesquisas para compreendermos a importância *in vivo* destas descobertas e hipóteses aqui levantadas. Este estudo constitui o primeiro relato dos possíveis papéis da poliadenilação no decaimento de diferentes moléculas truncadas de RNAs em *Leishmania*. A identificação das proteínas envolvidas nas vias de “turnover” destes RNAs, bem como a identificação do papel diferencial deste mecanismo entre os diferentes estágios evolutivos e entre distintas cepas do parasito, permitirão futuras análises *in vivo* deste novo mecanismo para regulação gênica diferencial em *Leishmania*.

Em conjunto, nossos resultados indicam a presença de mais um mecanismo de regulação pós-transcricional da expressão de RNAs em *Leishmania*. De fato, a detecção de fragmentos de RNAs poliadenilados não abundantes e a direta correlação entre a presença destas moléculas e a presença de uma via de degradação estimulada pela poliadenilação em muitos sistemas investigados até o momento, incluindo bactéria, *Archaea*, mitocôndrias e cloroplastos, leveduras, e no núcleo e citoplasma de células humanas, sugerem que um mecanismo similar esteja envolvido no decaimento de RNAs em *Leishmania*. Até o momento, não há nenhuma descrição de um exossoma funcional ou de poli(A) polimerases relacionadas com este processo, ampliando as possibilidades de pesquisas neste campo no futuro.

Em *Leishmania*, este é um campo aberto de pesquisa, pois nosso trabalho mostra pela primeira vez que a regulação da expressão gênica e da qualidade dos RNAs parece envolver também os fenômenos de poliadenilação 3' de rRNAs e a poliadenilação interna de rRNAs e mRNAs.

CONCLUSÕES E CONSIDERAÇÕES FINAIS

- Nossos resultados de RNA-Seq sugerem fortemente que todos os rRNAs podem ser poliadenilados na extremidade 3'-matura e na extremidade 3'-não processada das subunidades, similar à cauda poli(A)⁺ canônica encontrada nos mRNAs;
- A poliadenilação dos rRNAs parece ser um evento comum e amplamente disperso entre os dois subgêneros uma vez que este evento foi previamente observado em duas diferentes espécies [*L. (V.) braziliensis* e *L. (L.) infantum*], assim como por nós também observado em *L. (V.) braziliensis* e *L. (L.) amazonensis*;
- Nossos dados mostraram pela primeira vez que moléculas de rRNAs e mRNAs são poliadeniladas internamente no núcleo e no citoplasma de *Leishmania*;
- Estes eventos de poliadenilação transiente parecem estar diretamente relacionados com mecanismos de controle de qualidade de RNAs, similar ao mecanismo descrito em um número ainda limitado de organismos eucarióticos;
- A subunidade beta do LSU 28S rRNA poliadenilada na extremidade 3' ("canônica-like") apresentou um tamanho menor do que o esperado, sugerindo a presença de um sítio de poliadenilação alternativo anterior ao fim da região codificante para este rRNA;
- Nossos dados também sugerem que a poliadenilação interna de rRNAs e mRNAs pode ocorrer na fração polissomal de ribossomos ativos, sugerindo que rRNAs e mRNAs utilizados na tradução possam ser controladamente levados à degradação, provavelmente após seu uso;
- Identificamos em *L. (V.) braziliensis* sequências de genes ortólogos ao complexo do exossoma e ao complexo auxiliar (TRAMP) de humanos e leveduras, suportando a hipótese de que este fenômeno também está presente em *Leishmania*. Esta conclusão é também reforçada pela comprovação experimental da presença do exossoma em *T. brucei*, organismo taxonomicamente próximo de *Leishmania*.

O fenômeno de poliadenilação de RNAs truncados foi observado em moléculas de rRNAs e mRNAs e expande nossa visão a respeito da regulação pós-transcricional da expressão gênica em *Leishmania* e organismos relacionados. Isto adiciona mais complexidade ao paradigma de que quase todo o genoma em tripanosomatídeos é constitutivamente expresso, no qual a integridade das mensagens poderia realizar um papel crucial na regulação do conjunto de genes que são efetivamente traduzidos.

Estes dados ampliam novos campos de pesquisa em *Leishmania* e em Kinetoplastida em geral e destacam a importância de avaliar estas modificações pós-transcricionais e a estabilidade dos transcritos durante as análises de transcriptoma em escala genômica. Um estudo mais profundo do grupo de genes que são alvos de degradação direcionados pela adição de caudas poli(A)⁺ faz-se necessário e poderá auxiliar no melhor entendimento das adaptações biológicas destes parasitos em relação aos diferentes ambientes, destacando genes que poderiam ser inibidos numa determinada situação, influenciando as interações parasito-hospedeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

Afonso, L. C. e P. Scott. Immune responses associated with susceptibility of C57BL/10 mice to *Leishmania amazonensis*. Infect Immun, v.61, n.7, Jul, p.2952-9. 1993.

Alves, L. R., A. R. Avila, *et al.* Proteomic analysis reveals the dynamic association of proteins with translated mRNAs in *Trypanosoma cruzi*. Gene, v.452, n.2, Mar 1, p.72-8. 2010.

Barbazuk, W. B., S. J. Emrich, *et al.* SNP discovery via 454 transcriptome sequencing. Plant J, v.51, n.5, Sep, p.910-8. 2007.

Benz, C., D. Nilsson, *et al.* Messenger RNA processing sites in *Trypanosoma brucei*. Mol Biochem Parasitol, v.143, n.2, Oct, p.125-34. 2005.

Berriman, M., E. Ghedin, *et al.* The genome of the African trypanosome *Trypanosoma brucei*. Science, v.309, n.5733, Jul 15, p.416-22. 2005.

Bogdan, C., N. Donhauser, *et al.* Fibroblasts as host cells in latent leishmaniasis. J Exp Med, v.191, n.12, Jun 19, p.2121-30. 2000.

Britto, C., C. Ravel, *et al.* Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World *Leishmania* genomes. Gene, v.222, n.1, Nov 5, p.107-17. 1998.

Brown, K. M. e G. M. Gilmartin. A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. Mol Cell, v.12, n.6, Dec, p.1467-76. 2003.

Callebaut, I., D. Moshous, *et al.* Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. Nucleic Acids Res, v.30, n.16, Aug 15, p.3592-601. 2002.

Campbell, D. A., S. Thomas, *et al.* Transcription in kinetoplastid protozoa: why be normal? Microbes Infect, v.5, n.13, Nov, p.1231-40. 2003.

Chen, C. Y., R. Gherzi, *et al.* AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. Cell, v.107, n.4, Nov 16, p.451-64. 2001.

Clayton, C. e M. Shapira. Post-transcriptional regulation of gene expression in trypanosomes and leishmanias. Mol Biochem Parasitol, v.156, n.2, Dec, p.93-101. 2007.

Clayton, C. E. Life without transcriptional control? From fly to man and back again. EMBO J, v.21, n.8, Apr 15, p.1881-8. 2002.

Cloonan, N., A. R. Forrest, *et al.* Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods, v.5, n.7, Jul, p.613-9. 2008.

Colgan, D. F. e J. L. Manley. Mechanism and regulation of mRNA polyadenylation. Genes Dev, v.11, n.21, Nov 1, p.2755-66. 1997.

Croft, S. L. e V. Yardley. Chemotherapy of leishmaniasis. Curr Pharm Des, v.8, n.4, p.319-42. 2002.

De, A. J. [Prevention of kala-azar in Ceara, Brazil]. Rev Inst Med Trop Sao Paulo, v.3, Jul-Aug, p.175-80. 1961.

De Almeida Marques-Da-Silva, E., J. C. De Oliveira, *et al.* Extracellular nucleotide metabolism in *Leishmania*: influence of adenosine in the establishment of infection. Microbes Infect, v.10, n.8, Jul, p.850-7. 2008.

De Souza, M. C., E. A. De Assis, *et al.* The influence of ecto-nucleotidases on *Leishmania amazonensis* infection and immune response in C57B/6 mice. Acta Trop, v.115, n.3, Sep, p.262-9. 2010.

De Vries, H., U. Ruesegger, *et al.* Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. EMBO J, v.19, n.21, Nov 1, p.5895-904. 2000.

Deane, M. P. e L. M. Deane. [Experimental *Phlebotomus longipalpis* infection in a human case of visceral leishmaniasis]. Hospital (Rio J), v.46, n.5, Nov, p.487-9. 1954.

Decker, C. J. e B. Sollner-Webb. RNA editing involves indiscriminate U changes throughout precisely defined editing domains. Cell, v.61, n.6, Jun 15, p.1001-11. 1990.

Decuyper, S., J. Vandesompele, *et al.* Differential polyadenylation of ribosomal RNA during post-transcriptional processing in *Leishmania*. Parasitology, v.131, n.Pt 3, Sep, p.321-9. 2005.

Desjeux, P. Leishmaniasis: current situation and new perspectives. Comp Immunol Microbiol Infect Dis, v.27, n.5, Sep, p.305-18. 2004.

Deutscher, M. P. Degradation of RNA in bacteria: comparison of mRNA and stable RNA. Nucleic Acids Res, v.34, n.2, p.659-66. 2006.

Dheur, S., T. A. Vo Le, *et al.* Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation. EMBO J, v.22, n.11, Jun 2, p.2831-40. 2003.

Dong, L., R. V. Jensen, *et al.* Differentially expressed alternatively spliced genes in malignant pleural mesothelioma identified using massively parallel transcriptome sequencing. BMC Med Genet, v.10, p.149. 2009.

Edmond, M. B., S. E. Wallace, *et al.* Nosocomial bloodstream infections in United States hospitals: a three-year analysis. Clin Infect Dis, v.29, n.2, Aug, p.239-44. 1999.

Edmonds, M. A history of poly A sequences: from formation to factors to function. Prog Nucleic Acid Res Mol Biol, v.71, p.285-389. 2002.

Egecioglu, D. E., A. K. Henras, *et al.* Contributions of Trf4p- and Trf5p-dependent polyadenylation to the processing and degradative functions of the yeast nuclear exosome. RNA, v.12, n.1, Jan, p.26-32. 2006.

El-Sayed, N. M., E. Ghedin, *et al.* The sequence and analysis of *Trypanosoma brucei* chromosome II. Nucleic Acids Res, v.31, n.16, Aug 15, p.4856-63. 2003.

El-Sayed, N. M., P. J. Myler, *et al.* The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science, v.309, n.5733, Jul 15, p.409-15. 2005.

_____. Comparative genomics of trypanosomatid parasitic protozoa. Science, v.309, n.5733, Jul 15, p.404-9. 2005.

Emrich, S. J., W. B. Barbazuk, *et al.* Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res, v.17, n.1, Jan, p.69-73. 2007.

Estevez, A. M., T. Kempf, *et al.* The exosome of *Trypanosoma brucei*. EMBO J, v.20, n.14, Jul 16, p.3831-9. 2001.

Ewing, B. e P. Green. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res, v.8, n.3, Mar, p.186-94. 1998.

Ewing, B., L. Hillier, *et al.* Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res, v.8, n.3, Mar, p.175-85. 1998.

Fatica, A., M. Morlando, *et al.* Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3'-processing apparatus. EMBO J, v.19, n.22, Nov 15, p.6218-29. 2000.

Fenn, K. e K. R. Matthews. The cell biology of *Trypanosoma brucei* differentiation. Curr Opin Microbiol, v.10, n.6, Dec, p.539-46. 2007.

Fleischmann, J. e H. Liu. Polyadenylation of ribosomal RNA by *Candida albicans*. Gene, v.265, n.1-2, Mar 7, p.71-6. 2001.

Fleischmann, J., H. Liu, *et al.* Polyadenylation of ribosomal RNA by *Candida albicans* also involves the small subunit. BMC Mol Biol, v.5, Oct 4, p.17. 2004.

Francino, O., L. Altet, *et al.* Advantages of real-time PCR assay for diagnosis and monitoring of canine leishmaniosis. Vet Parasitol, v.137, n.3-4, Apr 30, p.214-21. 2006.

Fuentes, V., G. Barrera, *et al.* Functional Analysis of Sequence Motifs Involved in the Polyadenylation of *Trichomonas vaginalis* mRNAs. Eukaryot Cell, Mar 30. 2012.

Gilinger, G. e V. Bellofatto. Trypanosome spliced leader RNA genes contain the first identified RNA polymerase II gene promoter in these organisms. Nucleic Acids Res, v.29, n.7, Apr 1, p.1556-64. 2001.

Gourbal, B., N. Sonuc, *et al.* Drug uptake and modulation of drug resistance in *Leishmania* by an aquaglyceroporin. J Biol Chem, v.279, n.30, Jul 23, p.31010-7. 2004.

Gramiccia, M. e L. Gradoni. The current status of zoonotic leishmaniases and approaches to disease control. Int J Parasitol, v.35, n.11-12, Oct, p.1169-80. 2005.

Guerfali, F. Z., D. Laouini, *et al.* Simultaneous gene expression profiling in human macrophages infected with *Leishmania major* parasites using SAGE. BMC Genomics, v.9, p.238. 2008.

Haag, J., C. O'huigin, *et al.* The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Mol Biochem Parasitol, v.91, n.1, Mar 1, p.37-49. 1998.

- Haile, S., M. Cristodero, *et al.* The subcellular localisation of trypanosome RRP6 and its association with the exosome. Mol Biochem Parasitol, v.151, n.1, Jan, p.52-8. 2007.
- Haile, S. e B. Papadopoulou. Developmental regulation of gene expression in trypanosomatid parasitic protozoa. Curr Opin Microbiol, v.10, n.6, Dec, p.569-77. 2007.
- Hall, N., M. Berriman, *et al.* The DNA sequence of chromosome I of an African trypanosome: gene content, chromosome organisation, recombination and polymorphism. Nucleic Acids Res, v.31, n.16, Aug 15, p.4864-73. 2003.
- Haracska, L., R. E. Johnson, *et al.* Trf4 and Trf5 proteins of *Saccharomyces cerevisiae* exhibit poly(A) RNA polymerase activity but no DNA polymerase activity. Mol Cell Biol, v.25, n.22, Nov, p.10183-9. 2005.
- Heo, I., C. Joo, *et al.* TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell, v.138, n.4, Aug 21, p.696-708. 2009.
- Herwaldt, B. L. Leishmaniasis. Lancet, v.354, n.9185, Oct 2, p.1191-9. 1999.
- Hirose, Y. e J. L. Manley. RNA polymerase II is an essential mRNA polyadenylation factor. Nature, v.395, n.6697, Sep 3, p.93-6. 1998.
- Holzer, T. R., W. R. McMaster, *et al.* Expression profiling by whole-genome interspecies microarray hybridization reveals differential gene expression in procyclic promastigotes, lesion-derived amastigotes, and axenic amastigotes in *Leishmania mexicana*. Mol Biochem Parasitol, v.146, n.2, Apr, p.198-218. 2006.
- Ivens, A. C., C. S. Peacock, *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. Science, v.309, n.5733, Jul 15, p.436-42. 2005.
- Kao, C. Y. e L. K. Read. Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in *Trypanosoma brucei*. Mol Cell Biol, v.25, n.5, Mar, p.1634-44. 2005.
- Kaufmann, I., G. Martin, *et al.* Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. EMBO J, v.23, n.3, Feb 11, p.616-26. 2004.
- Kerr, S. F. Molecular trees of trypanosomes incongruent with fossil records of hosts. Mem Inst Oswaldo Cruz, v.101, n.1, Feb, p.25-30. 2006.
- Kima, P. E. The amastigote forms of *Leishmania* are experts at exploiting host cell processes to establish infection and persist. Int J Parasitol, v.37, n.10, Aug, p.1087-96. 2007.
- Kolev, N. G., J. B. Franklin, *et al.* The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. PLoS Pathog, v.6, n.9. 2010.
- Kuai, L., F. Fang, *et al.* Polyadenylation of rRNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A, v.101, n.23, Jun 8, p.8581-6. 2004.
- Kumar, S. e M. L. Blaxter. Comparing de novo assemblers for 454 transcriptome data. BMC Genomics, v.11, p.571. 2010.

Kushner, S. R. mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem. IUBMB Life, v.56, n.10, Oct, p.585-94. 2004.

Lacava, J., J. Houseley, *et al.* RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell, v.121, n.5, Jun 3, p.713-24. 2005.

Lao, K. Q., F. Tang, *et al.* mRNA-sequencing whole transcriptome analysis of a single cell on the SOLiD system. J Biomol Tech, v.20, n.5, Dec, p.266-71. 2009.

Laskay, T., G. Van Zandbergen, *et al.* Neutrophil granulocytes--Trojan horses for Leishmania major and other intracellular microbes? Trends Microbiol, v.11, n.5, May, p.210-4. 2003.

Lebowitz, J. H., H. Q. Smith, *et al.* Coupling of poly(A) site selection and trans-splicing in Leishmania. Genes Dev, v.7, n.6, Jun, p.996-1007. 1993.

Legendre, M., S. Audic, *et al.* mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res, v.20, n.5, May, p.664-74. 2010.

Leifso, K., G. Cohen-Freue, *et al.* Genomic and proteomic expression analysis of Leishmania promastigote and amastigote life stages: the Leishmania genome is constitutively expressed. Mol Biochem Parasitol, v.152, n.1, Mar, p.35-46. 2007.

Li, Q., Y. Zhao, *et al.* Comparison of the expression profiles of promastigotes and axenic amastigotes in Leishmania donovani using serial analysis of gene expression. Parasitol Res, v.103, n.4, Sep, p.821-8. 2008.

Liang, X. H., A. Haritan, *et al.* trans and cis splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot Cell, v.2, n.5, Oct, p.830-40. 2003.

Lipoldova, M. e P. Demant. Genetic susceptibility to infectious disease: lessons from mouse models of leishmaniasis. Nat Rev Genet, v.7, n.4, Apr, p.294-305. 2006.

Lukes, J., M. Jirku, *et al.* Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol, v.44, n.5, May, p.521-7. 1997.

Maioli, T. U., E. Takane, *et al.* Immune response induced by New World Leishmania species in C57BL/6 mice. Parasitol Res, v.94, n.3, Oct, p.207-12. 2004.

Mangus, D. A., M. C. Evans, *et al.* Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. Genome Biol, v.4, n.7, p.223. 2003.

Marsden, P. D. Mucosal leishmaniasis ("espundia" Escomel, 1911). Trans R Soc Trop Med Hyg, v.80, n.6, p.859-76. 1986.

Martinez-Calvillo, S., D. Nguyen, *et al.* Transcription initiation and termination on Leishmania major chromosome 3. Eukaryot Cell, v.3, n.2, Apr, p.506-17. 2004.

Martinez-Calvillo, S., S. Yan, *et al.* Transcription of Leishmania major Friedlin chromosome 1 initiates in both directions within a single region. Mol Cell, v.11, n.5, May, p.1291-9. 2003.

- Marzochi, M. C., S. G. Coutinho, *et al.* Canine visceral leishmaniasis in Rio de Janeiro, Brazil. Clinical, parasitological, therapeutical and epidemiological findings (1977-1983). Mem Inst Oswaldo Cruz, v.80, n.3, Jul-Sep, p.349-57. 1985.
- Minvielle-Sebastia, L., K. Beyer, *et al.* Control of cleavage site selection during mRNA 3' end formation by a yeast hnRNP. EMBO J, v.17, n.24, Dec 15, p.7454-68. 1998.
- Moe, K. T., J. W. Chung, *et al.* Sequence Information on SSRs and SNPs through Transcriptome Analysis of Mungbean. J Integr Plant Biol, Nov 18. 2011.
- Momen, H. e E. Cupolillo. Speculations on the origin and evolution of the genus Leishmania. Mem Inst Oswaldo Cruz, v.95, n.4, Jul-Aug, p.583-8. 2000.
- Morin, R., M. Bainbridge, *et al.* Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. Biotechniques, v.45, n.1, Jul, p.81-94. 2008.
- Morlando, M., P. Greco, *et al.* Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. Mol Cell Biol, v.22, n.5, Mar, p.1379-89. 2002.
- Mortazavi, A., B. A. Williams, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, v.5, n.7, Jul, p.621-8. 2008.
- Mullen, T. E. e W. F. Marzluff. Degradation of histone mRNA requires oligouridylation followed by decapping and simultaneous degradation of the mRNA both 5' to 3' and 3' to 5'. Genes Dev, v.22, n.1, Jan 1, p.50-65. 2008.
- Murray, H. W., J. D. Berman, *et al.* Advances in leishmaniasis. Lancet, v.366, n.9496, Oct 29-Nov 4, p.1561-77. 2005.
- Nagalakshmi, U., Z. Wang, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, v.320, n.5881, Jun 6, p.1344-9. 2008.
- Nedea, E., X. He, *et al.* Organization and function of APT, a subcomplex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and small nucleolar RNA 3'-ends. J Biol Chem, v.278, n.35, Aug 29, p.33000-10. 2003.
- Okoniewski, M. J. e C. J. Miller. Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics, v.7, p.276. 2006.
- Peacock, C. S., K. Seeger, *et al.* Comparative genomic analysis of three Leishmania species that cause diverse human disease. Nat Genet, v.39, n.7, Jul, p.839-47. 2007.
- Pearson, R. D. e A. Q. Sousa. Clinical spectrum of Leishmaniasis. Clin Infect Dis, v.22, n.1, Jan, p.1-13. 1996.
- Proudfoot, N. New perspectives on connecting messenger RNA 3' end formation to transcription. Curr Opin Cell Biol, v.16, n.3, Jun, p.272-8. 2004.
- Proudfoot, N. e J. O'sullivan. Polyadenylation: a tail of two complexes. Curr Biol, v.12, n.24, Dec 23, p.R855-7. 2002.

- Proudfoot, N. J. Ending the message: poly(A) signals then and now. Genes Dev, v.25, n.17, Sep 1, p.1770-82. 2011.
- Reis-Filho, J. S. Next-generation sequencing. Breast Cancer Res, v.11 Suppl 3, p.S12. 2009.
- Rochette, A., F. Raymond, *et al.* Genome-wide gene expression profiling analysis of *Leishmania major* and *Leishmania infantum* developmental stages reveals substantial differences between the two species. BMC Genomics, v.9, p.255. 2008.
- Royce, T. E., J. S. Rozowsky, *et al.* Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic Acids Res, v.35, n.15, p.e99. 2007.
- Ryan, C. M., K. T. Militello, *et al.* Polyadenylation regulates the stability of *Trypanosoma brucei* mitochondrial RNAs. J Biol Chem, v.278, n.35, Aug 29, p.32753-62. 2003.
- Ryan, K., O. Calvo, *et al.* Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. RNA, v.10, n.4, Apr, p.565-73. 2004.
- Sanger, F., S. Nicklen, *et al.* DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A, v.74, n.12, Dec, p.5463-7. 1977.
- Santos, R. F., M. A. Possa, *et al.* Influence of Ecto-nucleoside triphosphate diphosphohydrolase activity on *Trypanosoma cruzi* infectivity and virulence. PLoS Negl Trop Dis, v.3, n.3, p.e387. 2009.
- Saxena, A., T. Lahav, *et al.* Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. Mol Biochem Parasitol, v.152, n.1, Mar, p.53-65. 2007.
- Shcherbik, N., M. Wang, *et al.* Polyadenylation and degradation of incomplete RNA polymerase I transcripts in mammalian cells. EMBO Rep, v.11, n.2, Feb, p.106-11. 2010.
- Siegel, T. N., D. R. Hekstra, *et al.* Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. Nucleic Acids Res, v.38, n.15, Aug, p.4946-57. 2010.
- Skaar, D. A. e A. L. Greenleaf. The RNA polymerase II CTD kinase CTDK-I affects pre-mRNA 3' cleavage/polyadenylation through the processing component Pti1p. Mol Cell, v.10, n.6, Dec, p.1429-39. 2002.
- Slomovic, S., E. Fremder, *et al.* Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. Proc Natl Acad Sci U S A, v.107, n.16, Apr 20, p.7407-12. 2010.
- Slomovic, S., D. Laufer, *et al.* Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark. Mol Cell Biol, v.25, n.15, Aug, p.6427-35. 2005.
- _____. Polyadenylation of ribosomal RNA in human cells. Nucleic Acids Res, v.34, n.10, p.2966-75. 2006.

Slomovic, S., V. Portnoy, *et al.* Detection and characterization of polyadenylated RNA in Eukarya, Bacteria, Archaea, and organelles. Methods Enzymol, v.447, p.501-20. 2008.

Souza, A. E., P. J. Myler, *et al.* Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. Mol Cell Biol, v.12, n.5, May, p.2100-7. 1992.

Steinmetz, E. J. e D. A. Brow. Ssu72 protein mediates both poly(A)-coupled and poly(A)-independent termination of RNA polymerase II transcription. Mol Cell Biol, v.23, n.18, Sep, p.6339-49. 2003.

Stevens, J. R., H. A. Noyes, *et al.* The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. Parasitology, v.118 (Pt 1), Jan, p.107-16. 1999.

Tian, B. e J. H. Graber. Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip Rev RNA, v.3, n.3, May, p.385-96. 2012.

Torres, T. T., M. Metta, *et al.* Gene expression profiling by massively parallel sequencing. Genome Res, v.18, n.1, Jan, p.172-7. 2008.

Tyler, K. M. e D. M. Engman. The life cycle of *Trypanosoma cruzi* revisited. Int J Parasitol, v.31, n.5-6, May 1, p.472-81. 2001.

Van Dijk, E. L., G. Schilders, *et al.* Human cell growth requires a functional cytoplasmic exosome, which is involved in various mRNA decay pathways. RNA, v.13, n.7, Jul, p.1027-35. 2007.

Vanacova, S., J. Wolf, *et al.* A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol, v.3, n.6, Jun, p.e189. 2005.

Wahle, E. e U. Rügsegger. 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol Rev, v.23, n.3, Jun, p.277-95. 1999.

Walker, J., J. J. Vasquez, *et al.* Identification of developmentally-regulated proteins in *Leishmania panamensis* by proteome profiling of promastigotes and axenic amastigotes. Mol Biochem Parasitol, v.147, n.1, May, p.64-73. 2006.

Wang, S. M. Understanding SAGE data. Trends Genet, v.23, n.1, Jan, p.42-50. 2007.

Wang, Z., M. Gerstein, *et al.* RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, v.10, n.1, Jan, p.57-63. 2009.

Wilhelm, B. T. e J. R. Landry. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods, v.48, n.3, Jul, p.249-57. 2009.

Wilhelm, B. T., S. Marguerat, *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature, v.453, n.7199, Jun 26, p.1239-43. 2008.

Wilusz, C. J., M. Gao, *et al.* Poly(A)-binding proteins regulate both mRNA deadenylation and decapping in yeast cytoplasmic extracts. RNA, v.7, n.10, Oct, p.1416-24. 2001.

Wincker, P., C. Ravel, *et al.* The Leishmania genome comprises 36 chromosomes conserved across widely divergent human pathogenic species. Nucleic Acids Res, v.24, n.9, May 1, p.1688-94. 1996.

Worthey, E. A., S. Martinez-Calvillo, *et al.* Leishmania major chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. Nucleic Acids Res, v.31, n.14, Jul 15, p.4201-10. 2003.

Wright, A. D., S. Li, *et al.* Phylogenetic position of the kinetoplastids, *Cryptobia bullocki*, *Cryptobia catostomi*, and *Cryptobia salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. Mol Biochem Parasitol, v.99, n.1, Mar 15, p.69-76. 1999.

Wyers, F., M. Rougemaille, *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell, v.121, n.5, Jun 3, p.725-37. 2005.

Zhao, J., L. Hyman, *et al.* Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol Mol Biol Rev, v.63, n.2, Jun, p.405-45. 1999.

ANEXOS

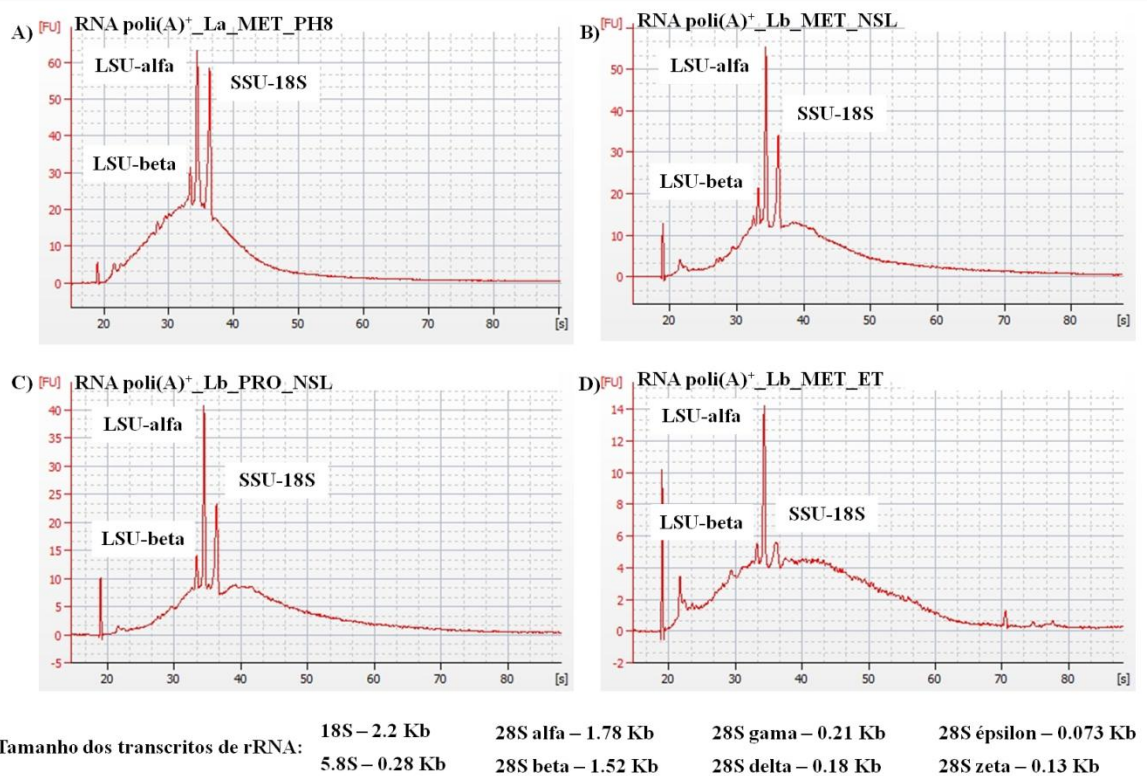


Figura suplementar S1. Padrão de eletroforese capilar das amostras de mRNAs poli(A)⁺ isoladas por afinidade oligo(dT) mostram a presença de subunidades de rRNAs em diferentes formas evolutivas de *L. (V.) braziliensis* e *L. (L.) amazonensis*. As duas formas evolutivas (promastigotas procíclicas e promastigotas metacíclicas) de todos os isolados apresentam rRNAs poli(A)⁺ co-purificados. Os perfis eletroforéticos dos mRNAs mostram a presença das subunidades de rRNAs mais abundantes. Todos os painéis representam os mRNAs poli(A)⁺ e os rRNAs poli(A)⁺ co-purificados após captura por beads magnéticos conjugados com oligo(dT). (A) RNAs poli(A)⁺ purificados das formas metacíclicas de *L. (L.) amazonensis* (isolado PH8). (B) RNAs poli(A)⁺ purificados das formas metacíclicas de *L. (V.) braziliensis* (isolado NSL). (C) RNAs poli(A)⁺ purificados das formas procíclicas de *L. (V.) braziliensis* (isolado ET). (D) RNAs poli(A)⁺ purificados das formas metacíclicas de *L. (V.) braziliensis* (isolado ET).

Texto Suplementar S1. Reads de rRNAs presentes nas bibliotecas transcriptômicas apresentam caudas homo- ou heteropoliméricas nas extremidades 3'-terminal. O conteúdo das seqüências e os respectivos scores de qualidade estão representados. Os resíduos de nucleotídeos sublinhados apresentam um score de qualidade Phred ≤ 10 .

Lbr_ET_MET

```
>GKLMG1101CH44D length=434 xy=0909_2523 region=1
run=R_2010_07_15_13_30_27_
GTAGAATTCAGAATTGTCCATAGGATAGCAAACCTCATCGGCGGGTTTTACCCAACGGTGG
GCCGCAATTCGGTCTGAATTCCTTCTCTGCGGGATTCCCTTTGTAATTGCACAAGGTGAAATTT
TGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCGACACGCGCACTACAATGT
CAGTGAGAACAAGAAAAACGACTTTGTGCGAACCTACTTGATCAAAGAGTGGGGAAACC
CCGGAATCACATAGACCCACTTGGGACCGAGGATTGCAATTATTGGTCGCGCAACGAGGA
ATGTCTCGTAGGCGCAGCTCATCAAACCTGTGCCGATTACGTCCCTGCCATTTGTACACAC
CGCCCGTCGTTGTTTCCAAAAAAAAAAAAAAAAAAAAAAAAACAAAAANAAAAAAAAAAAA
ACAAAAAAAAAAAAA
```

```
>GKLMG1101CH44D length=434 xy=0909_2523 region=1
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 30 27 27 27
27 37 37 37 37 37 37 37 37 37 39 40 39 37 37 33 33 30 32 30 28 28
24 24 26 26 21 21 23 23 30 30 37 35 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 34 34 28 28 28 28 37 37 34 34 34 34 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 32 32 30 30 30 32 35 32 32 25 25
25 33 33 28 30 30 30 27 30 30 34 34 34 37 37 37 30 30 30 34 37 37
37 37 37 37 37 37 37 35 33 32 32 29 29 30 29 23 15 14 12 12 12 12
22 27 17 27 19 19 19 19 33 26 35 35 35 35 35 33 35 24 21 21 20 33
33 28 24 21 21 19 19 29 20 27 27 19 19 19 19 23 23 23 19 19 19 19
14 14 15 15 25 25 28 21 21 22 21 22 13 15 15 15 21 19 21 21 21 21
21 25 24 27 22 23 19 19 19 19 23 27 27 32 32 32 32 32 33 33 25 25
24 32 32 32 32 27 23 23 23 27 28 28 28 28 27 27 24 23 24 23 27 30
30 31 23 23 24 29 31 31 31 31 32 31 31 30 27 19 19 19 19 25 24 27
27 33 33 33 33 35 35 35 35 35 35 33 30 30 35 35 35 35 30 25 25 25
30 25 35 35 35 35 35 35 35 35 28 28 28 29 30 17 23 23 23 23 28 28
28 33 33 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 21 7 7 7 7 7 7 7 0 8 10 10 10 10 10 10 10 10 10 7
```

```
>GKLMG1101C37RG length=244 xy=1161_0186 region=1
run=R_2010_07_15_13_30_27_
CTTAAGCCCAATTAGTGATTGCCATTATTCCTTTGGTAAGGGCGGAAACATACTCTTGG
AGAAAGTGTAGCCCTTCCGATGTAAGAATAGTAGGACCTAGGAGTTAGGACGAAGCTTA
TGGCGTGAGCCTAAGATGGACCGCCTCTAGTGCAGATCTTGGTTGGCGTAGCAAAGATC
TAACGGAGATATACTCAACATACAACGTTGGAAAAAAAAANAAAAANAAAAAAAAAAAAA
AAAN
```

```
>GKLMG1101C37RG length=244 xy=1161_0186 region=1
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 31
31 31 39 39 39 39 39 39 39 39 39 39 39 39 39 39 40 40 39 39 39 37
35 35 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 28 28 28 37
37 37 37 37 37 37 34 30 30 30 27 27 35 32 32 25 25 25 30 34 34 37
37 37 37 37 37 35 37 37 37 35 32 25 21 21 20 20 20 20 20 20 27 27
32 31 32 32 32 32 35 30 30 32 30 34 37 37 37 37 37 37 37 35 35 35
37 37 37 37 37 37 37 37 37 37 37 37 28 28 28 28 37 37 37 37 37 37
```

```

37 37 37 37 32 31 31 37 37 37 35 35 35 35 35 35 35 35 32 21 21 21
30 29 35 35 35 35 35 37 37 37 35 35 35 37 37 37 37 37 37 37 35
35 34 37 37 37 27 27 27 37 31 31 30 30 12 12 16 16 16 16 16 12 12
0 12 12 12 12 12 0 20 20 22 20 20 12 12 12 12 12 12 12 12 12 12
0

```

```

>GKLMG1101DYMLN length=444 xy=1507_1881 region=1
run=R_2010_07_15_13_30_27_
AAGAATTAGAGGCACCAAGTGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAA
CCGCTTAGGGATATGTGGCGTCAAACTTATGGGCTCATTCTCGATACGTGAAAAGGGTG
TTGGTGGATATGGACAGTTGGACGGTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGT
AACAACTCACCAACCGAATCCACCGGCCCGAAAATGGATGACGCTTAAGCCCAATTAGT
GATTGCCCATTTCTTTGGTAAGGGCGGAAACATACTCTTGGAGAAAGTGTAGCCCTT
CCGATGTAGATGTGGCCTGGAGGTTAGGACGAAGCTTATGGCGTGAGCCTAAGATGGACC
GGCCTCTAGTGCAGATACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAACAAAAAAAA

```

```

>GKLMG1101DYMLN length=444 xy=1507_1881 region=1
run=R_2010_07_15_13_30_27_
31 31 36 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 39 39
39 40 40 39 39 40 40 40 40 40 40 40 39 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35 37 30
30 24 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 35 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 30 28 28 28 37 30 30 28 28 35 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 30 30 30 38 38 38 35 35 30 30 23 32 35
30 30 30 35 23 23 18 18 18 30 25 32 32 35 35 35 35 38 30 30 31 35
35 40 40 40 38 38 38 38 38 38 35 35 34 34 25 25 37 32 32 32 35 35
35 35 35 35 35 35 35 32 32 32 35 35 35 35 35 35 35 35 30 30 30 32
32 32 32 35 35 35 35 35 35 35 35 35 35 35 35 35 40 40 35 35 35 35
35 35 40 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35
35 35 29 30 29 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 0
7 7 7 7 7 7 7 7 0 10 10 10 10 10 10 10 10 10 10 0 7 7
7 7 7 7 7 7 7 0 7 7 7 7 7 7 7 7 7 7 7 7 7
7 7 7 7

```

```

>GKLMG1101D0YL7 length=145 xy=1534_0173 region=1
run=R_2010_07_15_13_30_27_
AACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGCGTCGAGG
GGCGATTTACGTTCCCGACGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGAA
CAAAAAAAAAAAAAAAAAAAAAAAAA

```

```

>GKLMG1101D0YL7 length=145 xy=1534_0173 region=1
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 39 39 39 40 40
40 40 40 40 40 40 40 40 40 40 40 38 39 37 37 32 25 25 25 32 27 27 27
19 19 19 19 18 23 21 21 27 22 28 27 21 22 15 15 13 13 25 18 27 21
21 21 21 19 14 12 12 17 17 21 18 19 14 17 17 14 13 13 21 22 28 24
24 24 27 19 19 19 24 23 23 27 27 27 31 26 26 28 28 28 26 30 30 32
31 31 31 23 24 15 15 15 23 23 25 12 12 12 12 12 12 14 0 10 10 10
10 19 19 19 18 18 18 16 16 14 14 12 12

```

```
>GKLMG1101CMLJZ length=228 xy=0960_1549 region=1
run=R_2010_07_15_13_30_27_
TATACGTCCCTCTCCAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACT
TCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTC
ATGCTCAGGGAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAA
AAAAAAAAAAAAAAAAANAAAAAAAAANAAAAAAAAANAAAAAAAAANAAAAA
```

```
>GKLMG1101CMLJZ length=228 xy=0960_1549 region=1
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 30 30 30 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 35
35 35 35 37 37 37 37 37 31 31 31 31 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 33 33 33
33 33 30 21 21 21 25 25 25 35 35 35 35 35 33 35 35 30 28 28 33 32
30 14 12 12 11 11 11 11 11 10 10 10 10 10 10 11 11 11 11 11 11 11
15 15 15 15 15 15 15 21 21 21 21 21 13 12 10 10 10 10 10 10 10 10
0 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12 10 10 0 10 10 10
10 10 10 10 10 0 10 12 12 12 12 10 10 0 10 10 10 10 10 10 10
10 0 10 10 10 10 10 10
```

```
>GKLMG1102FZA86 length=321 xy=2335_1576 region=2
run=R_2010_07_15_13_30_27_
GCCAGCACCCGCGGTAATTCCAGCTCCAAAAGCGTATATTAATGCTGTTGCTGTTAAAGG
GTTTCGTAGTTGAACTGTGGGCTGTGCAGGTTTGTCCCTGGTCCCGTCCATGTCCGGAT
TTGGTGACCCAGGCCCTTGCAGCCCGTGAACATTCAAAGAAACAAGAACACGGGAGTGG
TTCCTTTCCTGATTTACGCATGTCATGCATGCCAGGGGGCGTCCGTGATTTTTTACTGTG
ACTAAAGAAGCGTGACTAAAGCAGTCATTCGTACTTGTAATTAAGTAAAGACATGGGAGT
AACAAGGAAAAAAAAAAGGCAG
```

```
>GKLMG1102FZA86 length=321 xy=2335_1576 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 38 38
38 38 38 34 34 21 21 21 21 23 37 38 38 38 38 39 38 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 30 32 18 18 18 27 37 37 21 21 30 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 31 31 31 31 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 22 22 22 22 35
20 32 32 29 30 30 30 30 16 16 16 14 14 14 27 27 12 25 27 27 27 29
32 26 26 28 28 19 19 25 28 32 32 30 30 32 32 26 23 23 27 27 24 22
22 22 22 12 12 12 22 11 12 12 12 12 18 11 12 12 12 12 11 11 18 20
20 16 16 16 13 18 17 18 11 11 11 12 11 18 20 22 25 11 11 12 12 11
11 11 11 11 11 11 11 11 16 16 22 12 12
```

```
>GKLMG1102GJ61H length=194 xy=2573_1043 region=2
run=R_2010_07_15_13_30_27_
GCTTCTATTTTCGTTGAAGAACGCAGTAAAGTGCATAAGTGGTATCAATTGCAGAATCA
TTCAATTACCGAATCTTTGAACGCAAACGGCGCATGGGAGAAGCTCTATTGTGTCATCCC
CGTGCATGCCATATTCTCAGTGTCGAACAAAAAACAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAA
```

```
>GKLMG1102GJ61H length=194 xy=2573_1043 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 39 39 39 39 40 40 40 40 39 39 39
40 39 39 39 40 39 39 39 39 39 40 40 40 39 39 33 33 33 39 40 40 40
39 39 39 39 39 39 39 40 37 37 37 37 37 37 37 37 28 28 28 37 37 37
35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35
35 37 37 35 35 35 35 24 24 24 24 24 19 19 19 24 27 29 27 23 24 22
21 19 19 19 18 18 27 28 25 20 20 22 23 16 12 12 11 11 11 11 11 11
12 11 11 11 11 11 11 11 11 11 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 0 12 12 12 12 12 12 12 12
```

```
>GKLMG1102IDFEX length=176 xy=3316_1207 region=2
run=R_2010_07_15_13_30_27_
GGGAGAGGAGAGTGAAGGTGTGTCGGCGGCAGTTGCTTTCAAGTAGCTGTTCCGACTGGCC
ATAACTGAAAAGGGGCAACAGAGAACCCTGGGATTATATTCCAAAAAGAAATTGCATGTGG
GCCTCCTTCTGAGTGAAGAAACAAAAAAAAAAAAAAAAANAAAAAAAAAGAAACGTCGT
```

```
>GKLMG1102IDFEX length=176 xy=3316_1207 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 35 35 33 33 33
30 21 21 21 21 21 21 26 26 35 35 35 35 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 35 33 32 32 32 16 16 14 14 14 24 12 12 13
28 29 32 33 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 32 21 21 20 20 12 12 12 12 12 12 12 12 12 12 12 12
12 12 10 10 0 10 12 12 12 12 10 10 13 10 10 10 22 24 25 24 12 12
```

```
>GKLMG1102F8T5P length=476 xy=2443_3611 region=2
run=R_2010_07_15_13_30_27_
GGCTTGTGGCGGCCAAGCGTTCATAGCGACGTCGCTTTTTGATCCTTCGATGTCGGCTCT
TCCTAACCTAGCGCCGAGAAAGACGCTAAGGGTTGGATTGTTACCCACTGACAGGGAAC
GTGAGCTGGGTTTAGACCGTCGTGAGACAGGTTGGTTTTACCCTACTTAGCTGGAAATTG
CGAGAAAAAGATTATCCGTGCAAAGCGGACGTCGGGTCGTTCTTGGGCCTGGCGAGCGGT
GGGTAGTACTTGGTACGCTCATCACCACACACAAACACTGTGTAGAGACATAGAAGTGAG
ATTGTGAAGGGATCTCGCAGGTATCGTGAGGGAAGTATGGGGTAGTACGAGAGGAACTCC
CATGCCGTGCCCTAGTTTCTGGGGTTTGTGCAACGGCAAGTGCCCCGAAGCCATCGACA
ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT
```

```
>GKLMG1102F8T5P length=476 xy=2443_3611 region=2
run=R_2010_07_15_13_30_27_
35 35 35 37 37 37 35 35 35 37 35 35 35 35 39 39 39 39 40 40 40 40
40 40 40 40 40 39 39 37 37 35 35 35 25 20 22 22 22 22 28 28 32 32
32 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 35 35 35 35 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35
35 37 37 37 37 37 30 21 20 20 24 33 33 33 33 33 32 20 20 21 30 30
30 37 37 37 37 32 32 34 37 37 37 37 37 34 23 23 25 31 37 37 37
37 32 32 32 32 37 37 37 37 37 35 29 27 25 25 33 33 33 29 29 19 19
19 19 19 25 28 28 28 24 12 12 14 14 16 23 18 31 28 28 27 27 32
33 32 25 21 21 18 33 28 28 27 32 32 32 30 27 23 23 23 30 27 22 19
19 19 19 19 28 28 28 35 35 35 35 35 35 37 37 37 37 35 35 35 26 26
26 37 37 37 37 37 37 37 37 37 37 40 40 38 38 38 34 34 32 32 32
29 27 27 23 23 23 23 18 13 13 13 21 17 21 22 22 24 24 27 29 27 27
```

```

19 20 20 32 32 33 33 33 33 32 30 30 30 35 35 25 25 25 28 33 33 31
26 28 26 23 24 25 25 32 30 30 30 29 29 30 31 31 27 27 27 27 24 23
23 23 23 23 19 29 29 30 19 19 19 19 19 24 14 23 21 22 21 24 24 24
17 17 17 19 23 23 23 23 23 27 27 29 31 31 27 29 27 23 23 23 23 23
18 18 14 14 14 23 22 19 20 19 19 23 17 17 12 14 29 27 25 23 23 23
23 17 14 14 14 18 27 19 21 22 21 17 19 19 27 31 31 29 29 22 13 11
11 11 13 16 11 11 11 11 11 11 11 11 11 11 11 11 11 0 7 7 7 7
7 7 7 7 0 7 7 7 7 7 7 7 7 7 7 7 0 7 7 7 7
7 7 7 7 7 0 7 7 7 7 7 7 7 11

```

```

>GKLMG1102HUX3F length=146 xy=3105_3209 region=2
run=R_2010_07_15_13_30_27_
CGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGG
TGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGC
TCAGGGACAAAAAAAAAAAAAAAAAAAAA

```

```

>GKLMG1102HUX3F length=146 xy=3105_3209 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 28 27 21 21 21 27 37 37 38 38 37 35 35
35 37 37 37 37 37 37 27 27 27 36 37 37 37 37 37 37 37 37 39 38
38 37 37 37 37 39 39 37 37 35 32 32 32 35 35 35 32 21 21 21 21 35
35 33 33 33 24 24 26 30 35 32 28 28 28 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 35 35 35 34 35 33 33 33 33 33 21 21
21 21 25 25 37 37 37 37 37 35 35 35 35 30 30 28 35 35 16 16 16 16
16 16 16 14 14 21 14 14 0 15 18 19 15 14

```

```

>GKLMG1102GZJPE length=170 xy=2748_0496 region=2
run=R_2010_07_15_13_30_27_
AGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATT
TACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGAACAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```

>GKLMG1102GZJPE length=170 xy=2748_0496 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 39 39 39 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 37
37 32 32 39 40 40 39 36 26 26 25 25 37 37 37 37 37 37 37 37 35
35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 28 28 28 37 37 37 37 37 37 37 37 37 30 28 27 29 27
28 12 11 11 11 11 11 11 11 11 11 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 12 12 12 12 10 10 0 11 11 11 11 11 0

```

```

>GKLMG1102FK9WN length=185 xy=2175_2005 region=2
run=R_2010_07_15_13_30_27_
AACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTA
CGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGAACAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
GAAAG

```

```

>GKLMG1102FK9WN length=185 xy=2175_2005 region=2
run=R_2010_07_15_13_30_27_
39 39 39 39 39 39 39 39 39 39 39 39 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 24 23 23
23 38 38 36 35 35 26 26 26 27 36 39 39 39 39 39 39 39 39 39 39

```

```

39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 38 38 38 38 38
31 27 27 27 31 31 31 36 36 36 36 36 36 38 38 32 31 29 30 35 12 12
12 12 12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 0 10 10 10 10
10 10 10 10 10 0 10 12 12 12 12 12 12 10 10 10 10 10 10 10 10
10 0 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 10 10 0 10 10
12 12 12 10 13 11 11 11 12

```

```

>GKLMG1102IXJB6 length=114 xy=3545_1428 region=2
run=R_2010_07_15_13_30_27_
TGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTC
CCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGACAAAAA

```

```

>GKLMG1102IXJB6 length=114 xy=3545_1428 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 30 30 30 37 37 37 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 39 39 39 39 37 39 39 37 37 31 31 31 31 37 37
35 37 30 30 31 28 27 31 28 30 34 37 37 37 37 37 37 31 28 28 28
37 37 37 37 37 37 37 37 37 37 37 32 31 31 37 37 37 35 25 25 27
34 34 30 32 32 33 33 32 32 32 30 19 19 19 19 22 12 11 11 11 11
11 11 11 11

```

```

>GKLMG1102HMWKE length=153 xy=3014_0716 region=2
run=R_2010_07_15_13_30_27_
TCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTG
GGGCTCGAGGGGCATTTACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTC
AGGGAACAAAAA

```

```

>GKLMG1102HMWKE length=153 xy=3014_0716 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 39 40 40 40 40 40 40 40 40 40
40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 35 35 35 35 37 37 35 35 35 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 32 25 26 26
35 35 37 37 37 37 37 37 37 37 37 37 30 28 33 33 32 14 14 14 14 14
14 14 12 15 15 15 15 15 15 11 13 13 13 13 13 15 15 21 21 15 15

```

```

>GKLMG1102IQS38 length=117 xy=3468_2902 region=2
run=R_2010_07_15_13_30_27_
AGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGCTCGAGGGGCATTT
ACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGACAAAAA

```

```

>GKLMG1102IQS38 length=117 xy=3468_2902 region=2
run=R_2010_07_15_13_30_27_
37 35 35 35 37 37 37 37 37 37 35 23 23 23 37 37 37 37 37 37 37
39 39 39 40 39 39 39 39 39 39 39 40 38 37 35 35 35 35 35 28 27 15
15 15 22 22 28 22 19 13 13 13 13 22 22 13 13 13 24 27 28 22 23 23
23 27 32 32 32 35 35 35 32 32 34 34 37 37 37 35 33 33 33 33 33
32 28 18 18 18 25 25 28 33 31 32 32 32 30 32 30 23 23 21 21 25 11
11 11 11 11 11 11

```

```

>GKLMG1102IR6X9 length=110 xy=3484_1951 region=2
run=R_2010_07_15_13_30_27_
GGCATGCTTCACTTCGGTGGGCGTCGAGGGCGATTACGTTCCCGACGGCGCTGAACCTTG

```

AGGCCTGAAATTTTCATGCTCAGGGACAAAAAANAAAAAAAAAAAAAAAAAAAA

```
>GKLMG1102IR6X9 length=110 xy=3484_1951 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 35 35 35 32 25 25 23 29 23 23 23 30
22 22 18 14 14 11 11 11 22 13 14 12 12 14 14 14 11 11 16 16 16 17
19 15 26 21 23 23 23 24 30 29 31 31 31 19 19 19 19 21 26 27 26 22
22 13 13 13 17 17 17 19 19 27 28 28 27 24 25 21 18 21 24 22 11 11
11 11 11 11 11 0 10 10 10 10 17 17 20 20 20 18 18 18 18 16 14 14
```

```
>GKLMG1102H6IYU length=189 xy=3237_2804 region=2
run=R_2010_07_15_13_30_27_
GAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTT
ACGTCCCAGAGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGAACAAAAAAAAAA
AAAAAAAAANAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAANAAAAAAAA
AAAAAAAAAN
```

```
>GKLMG1102H6IYU length=189 xy=3237_2804 region=2
run=R_2010_07_15_13_30_27_
37 30 30 30 37 37 37 37 37 37 37 39 39 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 39 31 28
28 28 37 37 37 37 37 28 28 28 28 37 37 35 35 35 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 32 31 31 31 31 37 37 37 37 37 30 31
32 37 37 37 37 37 37 37 37 37 37 37 37 37 35 33 28 23 23 15 13 14
12 11 11 11 11 11 11 11 11 11 10 10 10 10 10 10 10 10 10 0 12 12
12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 0 12 12 12 12 12 12 12 12 0 10 10 10 10 10 10 10 10
10 10 10 10
```

Lbr_NSL_MET

```
>GPST6G001A28QK length=452 xy=0329_3434 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGGGGATCTTATGGGCCGGCCTGCGGCAGGGTTTACCCTGTGTCAGCACC
GCGCCCGCTTTTACCAACTTACGTATCTTTTCTATTTCGGCCTTTACCGGCCACCCACGGG
AATATCCTCAGCACGTTTTTCTGTTTTTTCACGCGAAAGCTTTGAGGTTACAGTCTCAGGG
GGGAGTACGTTTCGCAAGAGTGAAACTTAAAGAAATTGACGGAATGGCACCACAAGACGTG
GAGCGTGCGGTTTAATTTGACTCAACACGGGGAACCTTACCAGATCCGGACAGGATGAGG
ATTGACAGATTGAGTGTCTTTCTCGATTCCCTGAATGGTGGTGCATGGCCGCTTTGGTC
GGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGACGAGATCCAAGCTGCCCAGTAGAA
TTCAGAATTGTCCATAGGATGAAAAAAAAAAC
```

```
>GPST6G001A28QK length=452 xy=0329_3434 region=1
run=R_2010_10_20_15_48_06_
37 37 37 33 33 33 33 33 33 33 33 18 18 16 16 30 30 30 19 19 19 28
23 23 25 21 21 18 18 32 34 39 39 40 40 40 40 40 39 39 39 39 39 39
40 39 39 39 40 40 40 40 37 37 37 37 37 37 30 30 30 37 37 21 21 21
21 29 28 31 30 28 34 35 35 37 37 37 37 37 37 37 37 37 37 37 35 21
21 21 21 30 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 36 35 35 37 37 37 37 37 37 35 31 31 31 37 37 37 37
37 35 32 22 22 22 19 27 16 19 14 14 14 14 14 12 25 28 28 30 30 30
27 28 30 34 34 37 37 37 37 34 30 30 30 30 35 35 32 32 32 32 25 25
25 17 17 14 14 14 15 30 17 19 15 15 15 15 15 19 23 23 29 29 28 30
28 27 27 23 26 26 27 19 19 16 16 14 19 22 22 22 32 31 27 28 28 34
34 37 37 37 34 34 34 37 37 37 37 37 34 34 34 30 30 30 30 30 30 33
29 27 27 27 28 33 32 32 21 21 21 30 30 28 23 23 35 38 37 37 35 40
40 38 38 38 31 31 31 31 31 31 30 35 35 38 38 35 35 35 35 35 35 35
35 35 35 32 24 20 20 20 25 28 32 32 33 33 33 28 28 28 28 28 27 30
32 29 26 27 27 21 18 13 13 13 18 11 11 11 16 16 16 18 18 16 17 21
21 21 21 22 19 19 19 27 27 23 23 23 27 29 27 27 27 17 17 17 17 23
17 13 13 13 17 17 19 26 26 26 26 26 27 31 31 23 23 23 23 23 23 25
22 22 19 13 13 12 12 21 22 21 21 21 21 24 23 21 19 19 17 23 23 25
27 27 27 26 26 26 26 26 24 21 18 18 13 13 11 11 11 18 22 22 21 21
15 13 15 15 15 15 15 13 13 11 11 11 11 11 11 15 15 13 11 11 11 15
11 11 11 11 11 11 11 11 11 11 11 15
```

```
>GPST6G001ASD22 length=353 xy=0206_0956 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGTTCGAGAACAAGAAAAACGACTTTTGTGCAACCTAACTTGACTCAAAAAG
AAGTAGGGGAAACCCCGGAATCACATAGACCCACTTGGACCGAGGATTGCAATTATTGGT
CGCGCCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAACGTGCCGATTACGTCCCTG
CCATTTGTACACACCCGCCCGTCGTTGTTTCCGATGATGGTGCAATACAGGTGATCGGACA
GGCGGTGTTTTATCCGCCCGAAAAGTTCACCGATATTTCTTCAATAGAGGAAGCAAAAAGT
CGTAACAAGGTAGCTGTAGGTGAACCTGCAGCTGGAAAAAAAAAAAAAAAAAAGGAA
```

```
>GPST6G001ASD22 length=353 xy=0206_0956 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 27 25 21 20 20 33 28 25 25 23 23
24 18 18 18 18 18 32 32 21 30 26 26 26 26 32 27 32 30 32 32 27 23
15 15 15 23 19 19 19 19 19 19 18 17 17 17 19 27 16 14 19 19 18 19
19 19 19 23 23 23 19 19 21 21 24 24 25 25 28 28 28 32 32 32 27 27
27 24 27 27 23 17 19 19 19 19 19 23 32 32 30 29 27 19 19 19 24 27
27 27 27 27 27 28 27 27 27 27 27 27 27 15 15 15 15 18 24 27 24
27 29 29 29 24 23 25 20 20 30 35 35 35 35 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 31
```

```

31 30 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 32 32 32
30 30 30 32 28 28 28 35 35 35 35 35 35 32 32 32 32 33 32 30 29 28
29 17 18 19 23 15 15 15 15 15 14 29 31 33 28 28 28 35 38 38 38 40
40 40 37 25 25 25 38 38 38 35 25 23 23 27 27 29 25 19 16 15 13 13
24 24 24 24 22 20 18 12 12 12 14 14 27 30 30 30 30 30 32 33 35 35
35 35 35 35 35 30 20 20 20 25 15 15 15 23 31 32 28 28 29 30 30 30
30 24 22 17 17 12 12 11 11 11 11 11 11 11 11 11 11 11 11 11 22 16
16

```

```

>GPST6G001ASH89 length=441 xy=0207_2267 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATTGATTTTTTACTGTGACTAAAGAAGCGTGACTAAAGCAGTCATTTCGACT
TGAATTAGAAAAGCATGGGATAACAAAGGAGCAGCCTCTAGGCTACCGTTTTCGGCTTTTTGT
TGGTTTTAAAGGTCTATTGGAGATTATGGAGCTGTGCGACAAGCGCTCTCCCATCGCAAC
TTCCGGTTCGGTGTGTGGCGCCTTTGGAGGGGTTTAGTGCGTCCGGTGC GGCTCCGGTTC
GTCCGGCCGTAACGCCTTTTTCAACTCACGGCCTCTAGGAATGAAGGAGGGTAGTTCCGGGG
GAGAACGTACTGGGGCGTCAGAGGTGAAATTCCTTAGACCGCACCAAGACGAACTACAGCG
AAGGCATTCCTCAAGGATACCTTCCTCAATCAAGAACCAAAAGTGTGGAGATCGAAGATGA
TTAGTAGCAAAAAAAAAAAAAA

```

```

>GPST6G001ASH89 length=441 xy=0207_2267 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 24 24 24 24 24 24 40 40
40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 39 39 40
40 40 39 39 40 40 40 40 37 37 37 37 37 37 37 37 37 37 30 30
30 37 37 37 37 37 37 37 37 37 37 37 35 35 35 37 28 28 28 35 35
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 30
35 35 35 32 16 16 15 15 27 21 21 27 27 15 15 16 16 28 28 28 24 25
20 32 34 37 37 37 31 31 32 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 31 31 32 26 26 26 26 30 30 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 40 40 35 35 31 31 35 35 35 40
40 40 40 40 40 38 38 38 40 40 35 35 30 30 30 30 37 37 38 38 38 35
35 38 38 38 38 38 30 30 30 30 26 28 30 28 28 37 26 35 35 35 35 40
35 35 35 30 30 30 35 40 40 35 35 35 35 35 35 35 35 40 40 33 33 35
32 32 32 35 35 35 35 35 35 35 35 35 35 35 25 25 25 30 35 35 35 35
35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35 35
30 30 29 28 23 23 23 29 29 33 33 35 35 35 35 35 35 28 28 28 35 35
35 35 28 28 28 35 35 35 35 35 35 35 35 35 35 35 35 35 33 33 27
27 25 25 25 16 12 12 12 19 17 11 11 11 11 11 11 11 11 11 11 11
11

```

```

>GPST6G001COWY2 length=434 xy=0986_3164 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATCGGCGAGCGAAGAGGGAACGAACCTCGTTGCCGAATCGGGTCTTACAGGG
CCTTGAGTTGTGGCACAAATATCATTTTCGTGGTTGGTGCAGGAGTGGCGAAATTC AAGCG
CAAAGCAAACCCGTTGCTGAATACAACCTTCATGTGAGTATTGAGCCAAAGAAGGTGTTA
GCCCATTGAGCCATAAACCTGAGCGCCTCATGAACTGTATTAAGCGAGTAGCACTGTTTG
GGAATGCAGTGTCAAGTTGGCAGTATTTGTCTGCCAAAGTTAAATACAGAGTAGGAAGA
CCGATAGCGAACAAGTAGCGTGAGCGAAAGTTTGAAAAGCACTTTGGAAGAGAGTGAAAT
AGAACCTGAAGTCGTGACAAAGACAATAGGAAGTACCTCCATTTCAACAGCCCAAGAAAC
AGAGAAAAAAAAAAAAA

```

```
>GPST6G001COWY2 length=434 xy=0986_3164 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 30 30 30 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 30 30 30 37 37 30 30 30 37 37 37
37 37 37 37 37 37 37 37 37 35 35 35 18 18 18 18 19 19 27 27 27 32
32 32 32 28 30 30 32 35 32 32 32 35 35 28 30 25 25 23 23 25 25 32
24 25 25 32 30 28 28 28 25 25 25 25 25 25 32 32 33 33 28 28 30 35
25 27 27 27 37 35 35 33 35 35 35 33 33 33 20 19 19 27 27 31 30 19
19 19 19 24 24 30 32 30 30 28 28 28 33 28 28 28 28 28 28 28 25 25
25 25 32 32 32 35 33 28 28 28 33 35 35 35 30 28 28 28 32 32 33 33
32 32 17 16 16 29 25 31 31 32 27 27 15 15 16 19 23 32 23 22 19 23
19 24 33 33 33 33 33 32 25 25 25 25 33 24 25 25 28 33 33 32 32 32
32 35 35 35 35 35 35 35 32 32 32 35 35 35 35 35 35 33 23 23 23 30
23 23 21 13 13 13 11 11 16 18 23 24 16 16 16 16 17 19 19 23 23 33
33 31 33 33 28 28 29 35 25 30 28 28 28 28 30 30 30 35 33 33 33 33
35 35 35 33 23 23 23 29 22 29 27 26 23 17 17 17 21 21 17 18 13 13
13 21 23 25 25 23 23 26 27 33 33 33 35 30 29 29 29 33 33 28 23 23
23 25 12 17 16 19 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
```

```
>GPST6G001BBTQP length=259 xy=0427_2495 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATCTGAATGTCAACGTGACGAGATTCACCGACGCGCGGGTAAACGGCGGGA
GTAACATGACTCTCTTAAGGTAGCCAAATGCCTCGTCTTCCAATTAGAGACGCGCATGA
ATGGATTAATGAGATTCCTCTGTCCCGAGTTACTATCTAGCGAAACGACAGTCAAGGGA
ACGGAATTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAANAAAAAA
```

```
>GPST6G001BBTQP length=259 xy=0427_2495 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 28 28 27 37 37 37 34 30 29 29 37 23 25 25 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 32 31 31 37 37 37 37 37 35 35 35 35 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 33 32 32 32
26 20 20 18 16 15 14 19 26 28 24 24 28 11 11 11 11 11 11 11 11 11 11
11 11 11 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
12 12 12 12 12 12 12 12 12 10 0 10 10 10 10 10 10 10 10 10 10
```

```
>GPST6G001C06JQ length=359 xy=1126_2004 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGCCGTTTGTGTTTTTTCGTTTTTCTTCGAGTACACACGTTGACATTTTTGCA
CAAAAATCCCAACTGCAGACCGTACTCATCACCGCATCAGTCCCCAAGCATAGAATGC
TCTGGTACATAGAGAAAGGTACACTCAGGGAAGTCGGCAAAAATGATGCGTAAGTTTCGCAA
GAAGCATTGGCTCTGAGGGCAAAGTCAGAGAAAACCGGAAGGCCAGTGTCTCTCGCAT
GCAGCCGTTCTTCAGATGTTGTTTTGCGCTCGCCGCTTAGCACGCTGATAGCGACTGTA
CTGTTCGTCTGTGAAGGCTACCTTCCCAGAAAAAAAAAAAAAAAAAAAAA
```

```
>GPST6G001C06JQ length=359 xy=1126_2004 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 36 36 36 37 37 37 35 33 31 32 24 23 23 19 14 12 12
12 12 12 22 18 13 15 17 17 28 15 15 20 25 28 33 35 35 35 35 35 33
31 27 23 23 23 27 27 21 12 11 11 11 11 19 25 28 28 11 11 11 12 14
31 28 28 28 31 33 30 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 35 34 32 32 30 30 28 28 21 17 19 12 12 12 12 12 17 17 28
28 28 28 28 27 19 20 20 25 25 25 25 32 32 32 32 31 28 25 25 25 25
25 25 25 26 26 29 23 25 21 25 32 32 32 32 32 28 30 30 37 37 37 34
25 25 26 37 32 32 32 28 37 37 37 37 37 37 37 37 36 36 35 37 37 37 37
37 37 37 37 37 34 32 32 32 25 19 19 19 19 24 24 23 23 24 24 23 23
26 27 16 16 16 24 29 28 28 28 28 24 17 17 13 13 13 13 13 13 18 18
16 16 11 11 13 22 28 30 28 28 23 23 23 32 34 37 37 37 37 37 37 37
37 37 37 37 37 26 26 26 37 37 35 27 27 27 27 28 27 24 18 13 13 13
13 15 24 27 27 29 28 29 29 32 35 32 34 34 34 34 38 37 37 35 35 35
35 35 30 30 30 35 20 20 20 25 35 35 35 35 33 33 28 28 28 28 33 28
27 27 27 19 17 17 26 26 18 12 12 12 12 16 16 16 16 16 18 16 16 12
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 12 8 8 8 8 8 8
8 8 8 8 7 7 7
```

```
>GPST6G001A9OGW length=364 xy=0403_0654 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATCGGACGTCGGGTCGTTCTTGGGCCTGGCGAGCGGTGGGTAGTACTTGGT
ACGCTCATCACCACACACAAACACTGTGTAGAGACATAGAAGTGAGATTGTGAAGGGATC
TCGCAGGTATCGTGAGGGAAGTATGGGGTAGTACGAGAGGAACTCCCATGCCGTGCCTCT
AGTTTCTGGGGTTTGTGCGAACGGCAAGTGCCCCGAAGCCATCGCACGGTGGTTCTCGGCT
GAACGCCCTAAGCCAGAAGCCAATCCCAAGACCAGATGCCCAACCAACAACAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAA
```

```
>GPST6G001A9OGW length=364 xy=0403_0654 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 31 31 31
40 40 40 40 40 40 39 39 39 39 39 40 40 40 40 40 40 40 40 40 40 40
40 40 39 39 39 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 36 36 36 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 26 26 26 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 32 31 31 35 35 37 31 32 32 32 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
28 17 17 17 31 31 30 19 19 19 30 31 33 35 35 18 18 18 37 29 29 38
35 35 35 30 30 40 12 12 12 12 12 12 12 12 12 12 12 12 12 12 8 8 8
8 8 8 8 8 8 8 8 0 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
10 10 10 10 10 10 10 8 8 0 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
10 10 10 10 10 7 0 7
```

```
>GPST6G001D57MS length=236 xy=1593_3474 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGGGAGAAGAGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCT
TTTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTC
ACTTCGGTGGGGCTCGAGGGCATTTCAGTCCCAGGGCGCTGAACCTTGAGGCCTGAAAT
TTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPST6G001D57MS length=236 xy=1593_3474 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 39 39 39 40 40 33 33 33 40 40 40
40 40 40 39 39 39 39 39 39 39 39 39 39 39 39 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 37 37 37 37 37 37 37 21 21 16 16 14 23 20
30 30 27 21 13 13 15 19 21 32 29 29 15 15 16 23 25 32 37 35 35 33
33 33 35 35 37 37 30 25 21 21 21 34 34 37 34 30 30 30 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35
37 37 37 37 37 32 33 33 33 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 35 33 32 32 32 32 32 32 25
18 18 18 28 30 28 33 33 33 33 33 33 37 37 30 30 30 32 35 12 12 12
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 0 10 10 10 10 10
10 12 12 12 12 12 12 12 10 10 10 10 10 10 10 0
```

```
>GPST6G001B1LX8 length=239 xy=0721_1230 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGGGAGAAGAGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCT
TTTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTC
ACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCAGGGCGCTGAACCTTGAGGCCTGAAAT
TTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGA
```

```
>GPST6G001B1LX8 length=239 xy=0721_1230 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 31 31 31 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 37 37 37 37 37 37 37 30 30 30 30 28 37 29
37 37 37 37 37 37 37 37 37 37 37 37 26 26 26 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 31 31 31 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35
37 37 37 37 37 35 35 35 35 37 37 37 37 37 37 37 37 37 37 31 31 31 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
21 21 21 37 37 37 37 37 37 37 35 33 33 30 19 19 17 18 25 12 12 11
11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 10 10 10 10 10 10 10
10 10 10 10 10 10 0 12 12 12 12 10 10 18 12 12 10 12 11
```

```
>GPST6G001A90XR length=242 xy=0407_0429 region=1
run=R_2010_10_20_15_48_06_
ACGACACGTATGGGAGAAGAGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCT
TTTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTC
ACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCAGGGCGCTGAACCTTGAGGCCTGAAAT
TTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGA
AG
```

```
>GPST6G001A90XR length=242 xy=0407_0429 region=1
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 31 31 31 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 37 37 35 32 33 33 33 32 32 18 21 21 21 21 25 24
35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35
37 37 37 37 37 35 35 35 35 37 36 21 21 21 31 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
21 21 21 35 35 35 37 37 35 35 35 37 37 35 21 21 21 31 32 12 12 12
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 0 12 12 12
12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 0 10 10 10 10 10 18
```

```
>GPST6G002GNP5P length=390 xy=2613_1947 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATCGTGGAGCGTGC GGTTTAAATTTGACCCAACACGGGGAAC TTTACCAGAT
CCGGACAGGATAGAGGATTGACAGATTGAGTGTTC TTTCTCGATTCCCTGAATGGTGGTG
CATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGACGAGATC
CAAGCTGCCAGTAGAATTCAGAATTGTCCATAGGATAGCAA ACTCATCGGCGGGTTTTTA
CCCAACGGTGGGCCGCATTCGGTCGAATTC TTTCTCTGCGGGATTCC TTTGTAATTGCACA
AGGTGAAATTTTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCGACACGCG
CACTACAATGTCAGTGAAAAAAAAGCAAAG
```

```
>GPST6G002GNP5P length=390 xy=2613_1947 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 37 39
39 38 38 37 37 37 37 37 31 31 31 37 37 37 37 37 37 37 37 31
31 30 30 30 30 30 34 34 34 35 37 37 35 35 35 37 37 37 37 37
37 37 37 28 28 28 35 37 37 37 37 37 37 37 37 37 37 37 37 35
35 35 35 37 37 37 37 37 37 37 37 37 37 37 35 34 34 34 34 37 37
34 34 34 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 28 32 32
32 37 37 23 30 30 28 28 32 29 30 31 30 23 19 16 16 16 27 22 22 22
29 29 25 29 23 25 25 25 25 25 32 32 33 33 28 28 28 30 30 37 37 37
37 37 35 35 35 35 35 35 35 35 21 21 21 25 30 30 28 32 33 33 33
32 33 33 32 32 32 34 30 30 25 25 25 30 35 35 35 35 35 30 30 30 30
37 37 37 37 37 37 34 35 35 35 35 35 37 37 37 31 30 30 30 37 37 37
37 37 21 17 21 20 20 28 26 26 27 27 22 23 19 19 19 27 27 29 29 30
31 29 29 28 28 30 28 28 33 33 33 33 32 32 30 20 20 20 25 34 35 35
30 30 30 33 29 35 35 32 28 25 24 24 23 23 28 23 23 24 23 19 19 19
19 19 19 19 23 23 23 24 23 23 23 23 27 27 27 27 29 31 32 32 30 30
30 30 30 30 30 30 29 28 28 27 23 23 24 24 23 23 24 23 19 19 19 24
22 30 30 30 29 31 27 27 27 27 29 29 23 23 23 27 25 25 26 25 25 25
25 23 12 12 12 12 12 12 12 12 21 22 11 11 11 18
```

```
>GPST6G002F9X2K length=452 xy=2456_2090 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGTCAAACAGACGCGCAAGGGAGGAGATATGGTTGCCAGTACTTCTCTCG
TACTGAAAGGGAGATGCAAAATGCATCTGTGGTGTGCACACGTCAAATCCGTTTGATTTG
CCCAACACCCGACCGGCCCTCCGTGGGTTGCGTTGGAGTGTGCCTGTTTGGACCCGAAAGG
TGGTGAAC TATGCCTGAACAAGGTGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCGA
TGCTGACGTGCAAATCGCTCGTATGATTTGGGTATCGGAGCGAAAGACTCATCGAACAC
CTAGTAGCTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTAGTAGAAGTATTTG
TGCGGTAA CATT TTAGGTAAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAANAAAAAAAAAAGAAAAAAAAAGGGAAAGGG
```

```
>GPST6G002F9X2K length=452 xy=2456_2090 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 35 37 37 39 39 38
38 38 34 30 27 31 31 31 30 27 24 24 37 37 38 37 37 37 37 38 39 39
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 35 35 35 37 23 23 21 21 32 32 33 33 35 35 35
35 37 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 35 35 28 28 28 35 37 37 37 37 35 29 30 30 30 34 34 37
35 35 33 28 28 28 28 27 27 33 32 19 19 19 27 27 25 31 27 23 23 23
25 32 32 35 35 35 35 35 37 37 37 37 35 35 35 36 36 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 34 27 27 27
27 34 35 35 37 37 37 37 35 35 21 21 21 30 28 28 28 30 30 30 26 17
17 17 21 21 32 32 32 32 32 32 32 32 37 37 37 37 37 37 35 37 37 37
```

```

37 37 37 30 30 30 30 32 35 30 30 30 35 32 30 30 30 34 37 35 32 32
32 35 30 30 30 20 20 20 19 25 32 35 35 35 32 32 32 32 28 28 30 35
19 25 25 33 24 24 24 30 29 30 28 27 18 18 13 13 13 27 27 24 24 25
22 13 13 13 13 27 23 22 22 23 22 22 16 12 12 11 11 11 11 11 13
13 18 20 20 21 22 17 18 19 27 28 27 27 27 29 29 29 27 24 21 17 17
16 20 22 20 11 11 11 18 12 15 18 11 11 11 11 11 19 18 11 11 11
11 11 11 11 11 11 15 15 11 11 11 11 11 11 11 11 11 11 11 11 11
11 11 0 7 7 7 7 7 7 7 7 7 7 7 0 7 7 10 10 10 10
10 7 7 7 7 0 7 7 7 7 7 7 7 7 7 7 15 7 7 7 7
7 7 11 11 11 11 7 7 7 8 8 8

```

```

>GPST6G002I4OM4 length=272 xy=3626_3118 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATAGTCGGCATCCGCTAAGGAGTGTGTAACAACCTCACCAACCGAATCCACC
GGCCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCCATTTCTTTGGTAA
GGGCGGAAACATACTCTTGGAGAAAGTGTAGCCCTTCCGATGTAGATGTGGCCTGGAGGT
TAGGACGAAGCTTATGGCGTGAGCCTAAGATGGACCGCCTCTAGTGCAGATCTTGGTTG
CGGTAGCAAAGATCTAACGAGAAAAAAAAAAAA

```

```

>GPST6G002I4OM4 length=272 xy=3626_3118 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 39 39 40 40 40 39 39 37 37 35 35
39 40 40 40 40 35 35 37 39 39 39 39 39 40 38 37 37 37 37 37 37
37 37 37 37 35 35 32 32 32 32 32 32 32 28 28 28 33 33 24 24 26 26
35 26 26 26 26 37 37 37 37 37 37 35 35 37 37 37 37 37 37 37 30 30
30 35 35 35 35 35 32 32 32 35 32 32 32 30 37 37 37 37 34 34 34 34
37 35 30 30 30 30 30 27 32 34 23 23 23 37 37 37 37 37 37 37 37
37 37 35 34 34 34 37 37 34 34 30 30 28 33 33 33 33 33 33 28 25 25
25 25 23 23 35 35 35 37 34 34 34 37 37 37 37 37 37 37 37 35 37 35
35 30 30 30 30 30 28 28 28 28 28 33 31 32 33 25 25 25 33 27 27 19
19 19 31 30 30 19 19 15 15 15 29 28 29 28 26 27 24 27 23 23 23 23
24 19 20 20 32 33 33 32 32 32 35 37 37 37 37 37 37 37 35 33 33 25
24 24 24 30 29 15 15 15 24 22 27 25 21 13 12 12 18 23 23 11 11 11
11 11 11 12 12 12 12 12

```

```

>GPST6G002GRQI3 length=470 xy=2659_0637 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGGCCTTCTTCGGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAAGG
GATATGTGGCGTCAAACTTATGGGCTCATTCTCGATACGTGAAAAGGGTGTGGTGGAT
ATGGACAGTTGGACGGTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACCTCA
CCAACCGAATCCACCGGCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCCA
TTATTCTTTGGTAAGGGCGGAAACATACTCTTGGAGAAAGTGTAGCCCTTCCGATGTAG
ATGTGGCCTGGAGGTTAGGACGAAGCTTATGGCGTGAGCCTAAGATGGACCGCCTCTAG
TGCAGATCTTGGTTGGCGTAGCAAAGATCTAACGGAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```

>GPST6G002GRQI3 length=470 xy=2659_0637 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 39 39 35 35 37 39 39 39 39 39 39
37 37 37 37 35 37 37 37 37 37 21 21 21 21 27 27 37 37 35 35 33 32
32 33 35 35 35 35 32 18 18 18 18 25 33 33 26 33 28 28 28 35 37 37
37 37 37 34 27 25 25 35 33 33 33 33 33 18 21 21 21 30 30 30 35
32 33 33 33 33 35 37 37 37 37 30 27 23 23 23 25 35 35 35 35 35
35 37 35 35 35 34 37 37 37 37 37 37 35 35 35 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 30 28

```

28	32	33	33	35	33	18	18	18	18	21	34	37	37	37	37	32	32	29	32	33	25
25	25	21	19	16	16	16	16	32	31	32	31	31	31	31	33	35	37	37	37	37	37
37	35	35	35	37	37	37	37	37	37	37	37	37	37	37	35	25	25	23	31	31	32
30	15	15	15	13	13	13	15	26	26	22	24	24	15	15	15	30	22	22	28	28	27
32	27	30	28	27	28	33	32	32	29	23	24	23	23	23	28	32	30	19	19	20	25
28	30	38	25	25	19	19	38	38	38	38	38	38	38	35	35	34	30	30	28	32	32
33	33	29	24	15	15	15	15	25	32	32	32	32	32	32	32	33	20	15	15	15	28
33	33	33	33	32	27	28	28	25	25	25	32	35	35	35	35	35	35	35	35	35	35
35	35	35	25	18	18	18	30	35	35	33	33	33	31	31	30	18	18	19	19	28	28
28	28	35	28	28	28	35	35	29	29	29	35	23	27	25	24	22	22	22	19	19	11
11	11	11	11	11	11	11	11	11	11	11	11	11	7	7	7	7	7	7	7	7	7
0	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
7	7	8	8	8	8	8	8														

```
>GPST6G002G0NZ5 length=487 xy=2760_3571 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGTCAAACAGACGCGCAAGGGAGGAGATATGGTTGCCAGTACTTCTCTCG
TACTGAAAGGGAGATGCAAAATGCATCTGTGGTGTGCACACGTCAAATCCGTTTTGATTTG
CCCAACACCGACCGGCCCTCCGTGGGTTGCGTTGGAGTGTGCCCTGTTTGGACCCGAAAAGG
TGGTGAACCTATGCCTGAACAAGGTGAAGCCCGGGGAAACCCAGGTGGAGGCCTGTAGCG
ATGCTGACGTGCAAATCGCTCGTATGATTTGGGTATCGGAGCTAAAGACTCATCGAACCA
CCTAGTAGCTGGTTCACATCGAAGTTCCCTCAGGATAGCTGGTGCTAGTAGAAGTATTG
TGCGGTAAAGCAAATGATTAGAGGCATTGGTGTCTTAGAATATCGACCTATTCTCAAAC
TTTAAATGTGCAAACAAACCGTAGCCTTAGCCAAGGACCGAAAGAAAAAAAAAA
TAACGTA
```

```
>GPST6G002G0NZ5 length=487 xy=2760_3571 region=2
run=R_2010_10_20_15_48_06_
39 39 39 39 39 39 39 39 39 39 39 40 40 40 21 21 21 38 40 40 40 40
40 40 40 40 40 40 30 30 30 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 21 21 21 21 36 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 29 24 24 24 24 37 37 39 39 39 39
39 39 39 39 26 26 29 36 39 39 33 33 33 39 39 39 39 39 39 36 36
36 39 39 37 33 32 28 33 29 28 20 11 11 11 11 11 11 11 11 16 16
16 17 17 23 24 26 33 38 38 39 39 39 39 36 36 36 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 35 35 35
39 35 35 21 21 21 34 34 39 36 38 38 38 38 38 39 39 39 26 26 26 26
35 36 38 38 38 38 38 38 26 26 26 26 31 31 31 38 36 35 33 33 33
36 36 21 21 21 21 33 38 38 31 31 31 38 38 36 33 33 33 39 33 33
27 39 33 33 33 39 39 39 39 39 39 39 39 33 33 33 39 39 39 39 39
26 26 21 21 18 24 24 24 28 30 32 31 33 33 19 19 19 31 32 29 29 39
39 40 35 35 39 39 35 35 35 28 28 28 28 28 33 33 31 31 18 18 18 23
30 32 29 28 28 27 25 26 29 32 32 26 24 24 17 17 17 27 23 24 24 24
24 21 17 17 17 17 17 17 17 26 16 15 17 17 22 18 22 12 12 12 12
24 17 12 12 12 17 22 22 21 22 22 21 22 22 19 25 25 22 21 21 17 17
11 11 11 17 17 13 11 11 11 11 11 11 11 11 11 11 11 11 11 11 19 20
16 13
```

```
>GPST6G002GQ02Y length=477 xy=2647_1256 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGTAAGCGGAGCAGGCGATGCGGAATGAACCGCTTAGGGATATGTGGCGT
```

CAAACTTATGGGCTCATTCTCGATACGTGAAAAGGGTGTGGTGGATATGGACAGTTGG
 ACGGTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACTCACCAACCGAATCC
 ACCGGCCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCATTATTCTTTGG
 TAAGGGCGGAAACATACTCTTGGAGAAAGTGTAGCCCTTCCGATGTAGATGTGGCCTGGA
 GGTTAGGACGAAGCTTATGGCGTGAGCCTAAGATGGACCGCCTCTAGTGCAGATCTTGG
 TTGGCGTAGCAAAGATCTAACGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AA

>GPST6G002GQO2Y length=477 xy=2647_1256 region=2
 run=R_2010_10_20_15_48_06_
 37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 37 37 37 40 39 38
 37 37 26 26 26 37 39 40 37 33 33 33 33 33 33 32 32 21 21 21 21 21
 30 35 26 33 28 30 30 37 37 37 37 37 37 37 37 35 35 33 33 35 33
 33 25 18 21 21 21 30 30 30 35 30 35 35 37 37 37 37 37 37 37 37
 37 31
 31 25 23 23 27 37 37 37 35 35 35 37 37 37 37 35 35 35 35 37 37
 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35
 35 37 37 37 32 30 30 35 35 28 28 28 26 32 21 21 21 21 37 37 37
 37 37 37 37 37 37 37 37 37 37 37 26 26 26 37 37 37 37 37 37 37
 37 37 37 37 37 37 37 37 37 34 30 30 30 32 32 26 26 26 31 29 22 23
 25 18 18 18 33 25 25 28 28 28 33 29 33 32 32 34 35 34 34 34 34 34
 35 38 38 38 38 38 38 38 38 38 38 38 25 25 19 19 37 37 37 37 38
 38 38 38 40 38 35 35 35 38 38 40 40 40 37 35 35 35 35 40 40 35 30
 30 23 25 25 35 35 25 25 25 32 32 32 35 35 40 35 35 35 40 40 40 35
 35 35 40 40 35 35 35 35 35 32 32 32 32 32 35 35 35 35 35 35 35
 35 35 35 35 35 35 35 35 23 23 23 23 18 18 18 18 35 33 28 28 28 33
 25 31 29 17 19 17 21 19 19 11 11 11 11 11 11 11 11 11 11 11 11
 11 11 10 10 10 10 10 10 10 7 7 7 7 7 7 7 7 7 7 7 7 7
 0 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 0 7
 7 7 7 7 7 7 7 7 0 7 7 7 7 7 7 7 7 7 7 0 7 7 7
 7 7 7 0 7 7 7 7 7 7 0 8 8 8 8

>GPST6G002H92OB length=470 xy=3278_0377 region=2
 run=R_2010_10_20_15_48_06_
 ACGACACGTATGCGTAGTCTTCGGCTATGCACCGTCTAGGTTTGGGAGTAATTTTATGG
 CGGTAACCCCGTCCCCACGGGAAAGGGTGGTGTGCTTTTTCTGCGTGCGTCTTCACGGGC
 GTTGCGCGGGCCCCGCTCTTTTTTATCACGTTTCCTTCGGGGGGCGTGACTCGTGGGTAC
 CCAACCGTGCTGTAAACTGAACCAACGGGATTGCCTCGTGCAGTTCGGGAGTCTTGTTTC
 GAGACATCTGCCAGATGGGGGATTTGGCTGGGGCGGCATATCTGTTACACGACAACGCAG
 GTGTCCCTAAGGCGAGCTCAGTGGGAACAGAAATCTCACGTAGAACAAGCGTAAAAGCT
 TGCTTGATTAACGATTTCCAGTACGAATCGAGACTGCGAAAGCAAAGGCCAGCGATCCTT
 TGACGAAGAATAGGAAAATATAGAAAAAAAAAAAAAAAAAAAAAAAAAAGT

>GPST6G002H92OB length=470 xy=3278_0377 region=2
 run=R_2010_10_20_15_48_06_
 37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 39 39 37
 26 26 26 39 39 39 39 39 37 37 37 37 39 39 39 39 39 39 39 39 39 39
 39 37 37 37 37 37 37 38 32 30 28 28 35 37 37 37 37 37 37 37 37
 31 32 32 32 37 35 24 24 24 24 37 37 37 37 37 37 37 37 31 31 32
 37 37 37 37 37 37 37 26 26 26 26 26 37 29 37 37 37 37 37 37 37
 37 34 34 34 34 30 21 21 21 34 32 37 37 37 37 37 37 37 37 32 32
 28 28 24 21 25 30 17 17 18 18 17 17 20 13 21 30 32 32 26 26 24 29
 29 25 24 18 14 14 14 14 14 12 27 14 22 23 24 24 24 27 27 27 26 26
 27 32 32 25 25 25 25 23 23 35 35 35 35 32 32 34 37 37 37 37 37
 37 30 30 30 31 37 37 37 37 21 21 21 37 37 37 37 37 37 37 35 35 35
 32 32 32 34 32 21 21 21 21 30 32 33 33 33 33 28 28 28 30 32 32 32

```

32 25 25 25 33 32 30 30 21 21 22 27 24 22 12 12 12 12 12 18 16 16
16 24 24 25 22 13 13 13 13 18 13 13 18 25 28 28 27 29 30 27 27 28
33 33 25 25 24 23 32 32 27 27 27 28 19 17 19 19 23 27 23 19 19 19
19 27 27 30 29 30 32 32 24 25 25 25 25 25 23 23 32 32 29 27 24 23
23 25 25 20 19 19 31 30 30 31 32 30 19 19 19 23 19 19 19 30 30 21
24 14 14 14 14 18 31 19 19 18 25 27 27 31 31 27 27 27 27 27 27 25
26 19 19 19 19 29 23 23 23 29 29 31 31 27 27 25 29 26 26 25 25 24
22 21 13 13 13 24 22 13 13 13 24 24 17 19 18 27 27 26 23 18 19 19
23 20 20 24 17 17 15 15 13 13 11 11 11 13 11 11 11 11 11 15 9
16 13 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 0 7 7 7
7 7 7 7 7 7 15 11

```

```

>GPST6G002IIJK1 length=464 xy=3374_2323 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGCGTAGTCTTCGGGCTATGCACCGTCTAGGTTTGGGAGTAATTTTATGG
CGGTAACCCCGTCCCCACGGGAAAGGGTGGTGTGCTTTTTTCTGCGTGCGTCTTCACGGGC
GTTGCGCGGGGCCCGCTCTTTTTTATCACGTTTCCTTCGGGGGGCGTGACTCGTGGGTAC
CCAACCGTGCTGTAAACTGAACCAACGGGATTGCCTCGTGCAGTTCCGGAGTCTTGTTTC
GAGACATCTGCCAGATGGGGAGTTGGCTGGGGCGGCATATCTGTTACACGACAACGCAG
GTGTCTTAAGGCGAGCTCAGTGGGAACAGAAATCTCACGTAGAACACAAGCGTAAAAGCT
TGCTTGATTAACGATTTCCAGTACGAATCGAGACTGCGAAAAGCAAGGCCAGCGATCCTTT
CGGAAGAAATAGGAAATATAGAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAN

```

```

>GPST6G002IIJK1 length=464 xy=3374_2323 region=2
run=R_2010_10_20_15_48_06_
31 31 32 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 39 39 39 37
21 21 21 37 37 39 39 39 39 39 37 37 39 40 40 40 40 40 40 39 39 39
39 39 39 40 40 40 40 40 35 35 35 35 37 37 37 36 30 28 28 28 31 37
31 31 32 32 37 35 26 26 26 26 37 37 37 37 37 37 35 35 23 23 23 37
32 32 33 33 33 33 32 21 23 23 23 23 32 24 37 37 37 37 37 35 35 35
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 28 21 21 18 22 20
20 20 22 20 22 20 11 12 12 12 12 12 19 15 27 27 27 27 21 21 21 27
30 25 24 18 14 14 14 14 14 14 23 14 28 28 30 30 30 32 32 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 34 34 34 34 32 32 35 35 35 35 37 37 30 30 30 30 37 37 37
37 37 34 32 32 32 32 33 33 31 32 32 30 27 15 15 16 15 24 13 17 17
21 26 26 28 25 15 15 15 15 27 18 20 27 27 27 27 28 24 23 19 20 20
32 32 32 32 32 33 35 35 35 35 35 32 28 27 28 33 29 27 27 19 15 15
15 23 23 29 28 30 30 30 30 32 28 30 30 33 33 33 35 35 35 35 35 33
33 33 35 25 25 25 32 33 32 32 30 27 15 15 15 19 23 23 23 27 27 31
31 21 21 21 21 31 33 28 28 28 28 23 19 18 18 28 28 33 33 33 35 35
29 28 28 28 28 30 35 35 23 23 23 35 35 35 30 29 35 35 30 23
23 23 23 23 29 35 35 35 35 35 33 30 30 30 33 30 29 29 23 19 19 17
17 16 17 17 19 21 21 23 23 20 21 19 21 21 18 18 22 21 12 16 16 17
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 0 7 7 7 7 7 7 7
7 0

```

```

>GPST6G002IL7M1 length=232 xy=3416_1435 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGTAGTACGAGAGGAACCTCCCATGCCGTGCCCTCTAGTTTCTGGGGTTTGT
CGAACGGCAAGTGCCCCGAAGCCATCGCACGGTGGTTCTCGGCTGAACGCCTCTAAGCCA
GAAGCCAATCCCAAGACCAGATGCCACCAACAACATAAATTTAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```
>GPST6G002IL7M1 length=232 xy=3416_1435 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 39 39 37 37 37
37 37 34 34 36 31 21 21 21 36 37 39 38 38 39 39 39 39 40 40 40 39
39 37 37 37 37 37 37 28 28 28 28 35 34 35 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 35 35 35 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 28 28 28 37 37 37 35 35 35 35 32
34 34 28 25 25 23 27 37 35 35 35 35 35 33 25 25 25 25 35 30 30 30
20 20 20 28 35 35 33 33 32 30 30 24 24 23 19 17 17 16 16 16 15 15
22 28 28 21 21 20 17 17 19 14 14 14 14 14 12 12 11 12 12 12 12
12 11 11 11 11 11 11 12 12 12 12 12 12 12 10 10 10 10 10 10 0
10 10 10 10 10 10 10 0 10 12 12 12 12 12 10 10 0 10 10 10 10
10 10 10 10 0 10 10 12 12 12 10 10
```

```
>GPST6G002F9KTJ length=160 xy=2452_1301 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATCGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATG
CTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCAGGCGCTGAACCTTGAGGCCTG
AAATTTTCATGCTCAGGGCAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPST6G002F9KTJ length=160 xy=2452_1301 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 39 39 39 40 40 40 40 40
39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40
40 40 40 39 39 39 39 39 37 37 37 37 37 37 37 37 37 37 37 37
35 31 27 27 25 32 21 22 22 22 30 24 19 19 19 15 15 15 13 25 28 18
18 17 18 19 19 19 23 19 19 19 24 28 29 30 29 30 24 19 16 16 18 18
28 27 27 30 29 32 33 33 35 35 30 34 34 34 34 37 37 37 34 32 32 30
30 29 21 21 19 28 11 11 11 11 11 11 11 11 11 11 12 12 12 12 12
12 15 18 18 12 12
```

```
>GPST6G002JT230 length=202 xy=3916_0430 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGG
CATTTACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANAAAAAAAAAAAAAAAAAAG
```

```
>GPST6G002JT230 length=202 xy=3916_0430 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 23 23 25 37 37 39 39
39 39 39 39 40 40 40 40 40 40 40 40 40 40 40 39 39 39 37 38 37
37 37 37 31 31 31 36 35 37 37 37 30 32 32 32 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 26 26 25 35 35 37 37 33 32 30 28 28 24 22 16
16 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 0 12 12 12 12
12 12 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 12
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 10 0 10 10
10 10 10 10 10 0 10 10 10 10 10 10 0 10 10 12 12 10 10 0
11 11 11 13
```

```
>GPST6G002G4PTN length=220 xy=2807_0041 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATCGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGG
GGCTCGAGGGGCATTTACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCA
```

GGGACAA
AAAG

>GPST6G002G4PTN length=220 xy=2807_0041 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 33 33 35 35 37 37 37 37
37 32 32 35 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37
37 37 37 37 37 37 35 32 30 28 17 16 16 28 25 11 11 11 11 11 11 11
11 11 11 11 11 11 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10
10
10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 0 10 10 10 13

>GPST6G002G8BFU length=119 xy=2848_0088 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGC
TGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGCAAAAAAAAAAAAAAAAAAAAAAAAAA

>GPST6G002G8BFU length=119 xy=2848_0088 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 37 37 37 39 39 37 37 37 37
34 34 34 37 37 37 28 26 26 26 33 33 32 30 29 19 19 19 19 27 27 15
15 15 19 19 30 24 23 23 23 33 33 35 37 37 37 37 37 37 37 37 37 37
30 30 30 30 33 32 32 29 24 24 23 23 23 16 16 15 24 27 30 28 28 28
27 25 13 13 13 22 11 11 11 11 11 11 11 11 11 11 11 10 10 10 10 10
10 10 10 0 14 14 14 10 10

>GPST6G002I4NH2 length=230 xy=3626_1640 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATTTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGC
GGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGA
GGCCTGAAATTTTCATGCTCAAGGAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAN

>GPST6G002I4NH2 length=230 xy=3626_1640 region=2
run=R_2010_10_20_15_48_06_
36 35 37 37 37 37 37 37 37 37 37 21 21 21 18 30 28 35 33 33 31 16 16
18 32 37 37 37 37 31 31 31 39 40 40 40 40 40 40 40 40 40 40 40 40
40 40 39 39 39 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 28 28 28 28 37 37 35 32 32 26
26 26 28 37 37 23 23 23 35 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 35 33 25 23 23 32 32 33 32 18 16 16 19 19 19
32 32 30 28 28 28 28 28 28 23 19 17 13 13 13 14 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10
12
12 12 10 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12 12 12 10 10
10 10 10 10 10 10 10 10 0

>GPST6G002FMXGZ length=215 xy=2194_1377 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATTTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGC
GGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGA

GGCCTGAAATTTTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAANAAAN

>GPST6G002FMXGZ length=215 xy=2194_1377 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 33 37 28 28 30 40 40 40 40 39 39
39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 39 39 39 40 40 40 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 35 35 35 35 37 37 37 37 28
28 28 28 37 37 28 28 28 37 37 37 37 37 37 37 37 37 37 37 37
37 37 36 36 36 36 37 37 37 37 37 37 37 37 35 21 21 21 35 35 37
37 37 37 37 37 37 35 30 28 27 30 28 12 12 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 10
10
10 10 10 10 10 10 10 14 14 14 14 12 0 11 11 11 0

>GPST6G002GZB77 length=169 xy=2745_3093 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCT
CGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGA
ACAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAAAAAAAAAAAAAAAAA

>GPST6G002GZB77 length=169 xy=2745_3093 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 34 25 21 21 21 25 25 25 25 32 33
30 30 30 37 37 38 38 40 39 39 39 39 39 39 39 39 39 39 39 37 37
37 35 33 33 32 32 28 28 28 28 18 18 18 18 33 33 32 30 30 22 22 19
19 30 24 16 16 16 16 24 27 24 23 23 23 28 28 29 29 27 27 27 24 19
15 15 15 15 27 24 27 28 27 27 24 24 27 29 15 15 15 19 19 17 28 25
25 25 28 25 20 18 11 11 11 12 12 18 11 11 11 11 11 11 11 11 11
11 11 10 10 10 10 10 10 10 10 10 10 0 10 12 12 12 12 12 12 12
12 12 10 10 0 10 12 12 12 12 12 12 12 12 10 10

>GPST6G002I3B5R length=213 xy=3611_1725 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGG
TGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGC
TCAGGGACAA
AAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAANAAAN

>GPST6G002I3B5R length=213 xy=3611_1725 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 28 28 28 35 39 39 39 39 37 23 23
25 37 39 40 39 39 39 39 39 39 40 40 40 40 40 40 39 39 39 39 39
39 37 37 39 39 37 35 35 37 37 37 37 37 37 37 32 32 32 32 37
37 37 35 32 18 18 18 18 28 32 16 16 16 30 25 25 23 37 37 37 35 33
35 35 35 35 37 37 37 21 21 21 21 27 25 33 33 35 35 32 32 34 37 21
21 21 37 37 37 37 37 37 37 37 37 37 30 30 28 33 32 12 12 12 12
12 12 11 12 12 12 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12
12
12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 0 14 14 14 14 14 0 11 11 11 0 11 11 0

```
>GPST6G002HCSP1 length=174 xy=2899_0215 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGC
TCGAGGGGCATTTACGTCCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGG
GCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAGAAG
```

```
>GPST6G002HCSP1 length=174 xy=2899_0215 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 39 40 40 40 40 39
39 39 39 39 39 39 40 40 40 40 40 40 40 39 39 39 39 39 39 39 39 39 39
37 37 37 37 36 33 33 33 37 37 37 31 31 31 28 37 37 33 33 32 21 21
19 19 28 22 11 11 11 14 19 22 13 13 13 13 13 18 30 30 30 30 29 30
31 33 32 32 21 21 21 21 35 35 32 32 25 25 25 33 30 30 34 21 21 21
34 37 37 37 35 33 33 32 27 26 26 30 12 12 12 12 11 11 11 11 11 11
11 11 10 10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 10 10 10
10 0 10 10 10 10 10 10 10 10 19 14 14 15 15 14 18 12 12 14
```

```
>GPST6G002JKMMI length=206 xy=3808_1528 region=2
run=R_2010_10_20_15_48_06_
ACGACACGTATGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGG
GCTCGAGGGGCATTTACGTCCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAG
GGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPST6G002JKMMI length=206 xy=3808_1528 region=2
run=R_2010_10_20_15_48_06_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 30 30 30 30 37 37 37 37 37
25 25 26 26 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 36 36 36 36 37 37 37 37 37 37 37 37 37 32 26 26 26 31 31
31 37 37 37 37 37 35 35 33 28 27 24 30 28 12 12 12 12 12 12 12 12
12 12 11 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 0 10 10 10 10 10 10 10 0 10 12 12 12 12 12 12
10 0 10 10 10 10 10 10
```

Lbr_ET_PRO

```
>GPUPS7E01B33ZO length=245 xy=0749_3234 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATTTTCTATTTTCGTTGAAGAACGCAGTAAAGTGCATAAGTGGTATCAATT
GCAGAATCATTCAATTACCGAATCTTTGAACGCAAACGGCGCATGGGAGAAGCTCTATTG
TGTCATCCCCGTGCATGCCATATCTCAGTGTCTGAACAAAAAACAAAAAANAAA
AAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAANAAAAAAN
AAAAAN
```

```
>GPUPS7E01B33ZO length=245 xy=0749_3234 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 39 39 40 40 40 40 39 39 39 40 40
40 40 40 40 40 40 40 40 40 40 40 40 39 33 28 28 28 39 39 39 39 39
37 37 37 37 39 39 33 33 31 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 25 26 26 31 35 37 37 37 37 37 37 37 35 35 36 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 32 32 32
32 29 28 14 14 14 14 14 14 19 12 12 12 12 12 12 12 12 12 12 12
0 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 10
10 10 12 12 12 12 12 12 12 12 0 10 10 10 10 10 10 10 0 12 12 12
10 10 0
```

```
>GPUPS7E01DQJYH length=196 xy=1415_2039 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATCAGACGCGCAAGGGAGGAGATATGGTTGCCAGTACTTCTCTCGTACTGA
AAGGGAGATGCAAAATGCATCTGTGGTGTGCACACGTCAAAATCCGTTTGATTTGCCCAAC
ACCGACCTCTCCTTCTCGTGCACGTCTGCCGCGCATGCACACGCGCGGTTGACCGTAA
AAAAAAAAAAAAAAAA
```

```
>GPUPS7E01DQJYH length=196 xy=1415_2039 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 33 33 33 33 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 35 35 35 37 37 37 37 35 35 35 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 33 31 19 19 17
24 18 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 12
```

```
>GPUPS7E01DBYAX length=516 xy=1249_0727 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGTGTCTTAGAATATCGACCTATTCTCAAACCTTAAATGTGCAAACAA
ACCGTGCCTTAGCCAACCTGCAAGGCCGAAAGAAAAGAATAGAGGCACCAAGTGGGCCT
TCTTCGGTAAGCGGAGCAGGCGATGCCGAATGAACCGCTTAGGGATATGTGGCGTCAAAA
CTTATGGGCTCATTCTCGATACGTGAAAAGGGTGTGGTGGATATGGACAGTTGGACGGT
GGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACCTACCAACCGAATCCACCGG
CCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCCATTTATTCCTTTGGTAAGG
GCGGAAACATACTCTTGGAGAAAGTGTAGCCCTTCCGATGTAGATGTGGCCTGGAGGTTA
GGACGAAGCTTATGGCGTGAGCCTAAGATGGACCGCTCTAGTGCAGATCTTGGTTGGCG
TAGCAAAGATACTAAACGGAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01DBYAX length=516 xy=1249_0727 region=1
run=R_2010_10_21_16_08_57_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 26 26 26 40 31
31 29 35 35 39 40 40 40 40 40 39 39 39 40 36 36 36 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 39 39 40 31 31 31 31 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40
34 34 34 34 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 39 27 20 20 20 20 32 32 33 38 30 35 35 36 37 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 39 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 35 35 37 40 40 40 38 36 36 32 28 27 27 11 11 11 11 11 11 11 11
11 22 20 24 32 32 35 35 35 36 33 35 35 38 38 38 27 21 21 21 21 27
27 38 38 38 36 36 38 38 38 21 21 21 36 39 39 35 32 32 32 32 36 39
39 40 40 30 33 33 36 36 36 36 28 28 26 26 26 25 19 26 25 35 34 34
32 32 33 33 33 35 20 20 20 36 36 35 35 35 35 33 33 33 31 31 23 23
33 33 34 35 35 35 35 35 30 30 29 29 26 26 26 32 32 26 23 23 19
19 16 24 24 25 23 15 17 17 21 27 26 26 29 24 28 28 28 24 28 25 25
24 25 25 23 23 23 18 15 18 23 26 26 23 23 13 13 13 18 17 17 18 28
25 25 25 18 18 17 28 27 25 25 28 28 19 19 15 15 18 19 25 25 21 22
20 20 20 24 15 24 17 20 11 11 11 11 14 15 18 11 11 11 11 11 11 11
11 11 11 11 11 11 11 11 11 11
```

```
>GPUPS7E01AJO4T length=520 xy=0107_0875 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATCTCAAACCTTTAAATGTGCAAACAAACCGTGCCTTAGCCAACTGCAAGGC
CGGAAAGAAAAGAATTAGAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAGCAGGCGATG
CGGAATGAACCGCTTAGGGATATGTGGCGTCAAACCTTATGGGCTCATTCTCGATACGTG
AAAAGGGTGTGGTGGATATGGACAGTTGGACGGTGGCCATGGAAGTCGGCATCCGCTAA
GGAGTGTGTAACAACCTCACCAACCGAATCCACCGGCCCGAAAATGGATGACGCTTAAGC
CCAATTAGTGATTGCCATTATTCCTTTGGTAAGGGCGGAAACATACTCTTGGAGAAAAGT
GTAGCCCTTCCGATGTAGATGTGGCTGGAGGTTAGGACGAAGCTTATGGCGTGAGCCTA
AGATGGACCGGCCTCTAGTGCAGATCTTGTTGGCGTAGCAAAAAGAAAAAAAAAAAAAAAA
ANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01AJO4T length=520 xy=0107_0875 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 39 39 39 40 39 31 31 17
17 17 29 33 33 33 18 16 18 25 37 37 37 37 37 37 37 37 39 39 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 30 30 30 30 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
32 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 28 28 26 26 21 21 21 37 37 32 32 30 30 30 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 35 35 34 37 35 35 35 35 25 20 20 20 28 25 25 25 32 32
23 23 25 32 32 33 33 33 33 33 32 32 32 35 32 30 30 30 34 34 35 35
34 34 32 33 33 33 32 29 27 27 15 15 15 15 27 15 15 15 15 24 27
32 33 33 33 35 35 38 40 38 38 38 35 37 35 35 40 40 35 35 35 35
35 35 32 32 32 32 15 15 15 24 27 29 30 29 29 27 27 13 13 13 27 27
22 29 29 18 18 18 19 15 15 23 23 23 21 19 27 30 24 23 23 19 19 19
```

19	19	23	26	26	26	27	29	23	23	23	23	26	23	23	19	19	14	14	27	27	27
27	27	29	27	27	31	29	29	27	27	27	27	29	23	23	23	23	23	23	23	27	25
25	27	27	25	29	29	28	28	28	28	35	29	30	30	35	35	35	35	35	35	35	35
35	17	17	17	23	25	25	25	25	17	17	15	15	23	19	21	19	20	21	21	22	23
21	21	21	23	25	25	19	19	12	12	12	12	23	23	12	12	12	12	19	25	21	21
21	21	21	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	0	7	7
7	7	7	7	7	7	7	7	7	0	7	7	7	7	7	7	7	15	8	8	8	8
0	10	10	10	10	10	0	9	9	9	9	9	9	9	9	9	9					

```
>GPUPS7E01DTIZS length=350 xy=1449_1494 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGTGGGAACAGAAATCTCACGTAGAACACAAGCGTAAAAGCTTGCTTGAT
TAACGATTTCCAGTACGAATCGAGACTGCGAAAGCAAGGCCAGCGATCCTTTGCGAAGAA
TAGGAAATATGAACAAATCATACCAGAGGTGTCAGAATAATTACCACAGGGATAACTGGC
TTGTGGCGGCCAAGCGTTCATAGCGACGTCGCTTTTTTGATCCTTCGATGTCGGACTAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01DTIZS length=350 xy=1449_1494 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 39 39 39 40 40 40 40 40 39
39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 31 31 31 31 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 31 31 31 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 36 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 31 31 32
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
11 11 11 11 11 11 11 11 11 0 12 12 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 10 10 8 8 8 8 0
8 10 10 10 10 10 10 10 8 8 0 8 8 8 8 8 8 8 8 8 8 8 8 8 8 12
12 12 12 8 8 0 8 8 12 12 12 8 8 0 8 8 8 8 8 8 8 8 8 8 8 8
```

```
>GPUPS7E01CY0WT length=537 xy=1101_3787 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATCGCTTTTTGATCCTTCGATGTCGGCTCTTCCTAACCTAGCGCCGAGAA
GACGCTAAGGGTTGGATTGTTACCCACTGACAGGGAACGTGAGCTGGGTTTAGACCGTC
GTGAGACAGGTTGGTTTTACCCTACTTAGCTGGAAATTGCGAGAAAAAGATTATCCGTGC
AAAGCGGACGTCGGGTCGTTCTTGGGCCTGGCGAGCGGTGGGTAGTACTTGGTACGCTCA
TCACCACACAAAACACTGTGTAGAGACATAGAAGTGAGATTGTGAAGGGATCTCGCAGG
TATCGTGAGGGAAGTATGGGGTAGTACGAGAGGAACTCCCATGCCGTGCCTCTAGTTTCT
GGGTTTGTGCGAACGGCAAGTGCCCCGAAGCCATCGCACGGTGGTTCTCGGCTGAACGCC
TCTAAGCCAGAAGCCAATCCCAAGACCAGATGCCACCAACAACATAATTTAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01CY0WT length=537 xy=1101_3787 region=1
run=R_2010_10_21_16_08_57_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 30 30 30 30 30 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 37
37 37 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 26 25 25 40 40
40 40 40 40 37 21 21 21 40 40 40 40 40 40 40 40 40 40 39 39 39 39
```

39	39	40	40	40	40	40	40	40	40	40	40	40	40	40	37	37	37	40	40	40	40	40	
40	40	34	34	34	34	40	39	39	39	40	40	40	40	40	40	40	40	40	40	40	37	37	33
33	33	40	40	40	40	40	40	40	26	26	28	28	26	39	27	40	40	40	40	40	40	40	40
40	40	40	40	39	39	39	40	37	37	37	40	40	40	40	37	29	29	29	37	40	40	40	40
40	40	40	37	37	21	21	21	37	37	40	40	40	40	40	40	40	40	29	29	29	34	34	34
34	34	37	38	32	28	28	28	35	40	40	39	39	40	40	40	40	40	40	40	40	40	40	40
40	40	40	40	40	40	40	40	40	39	39	39	40	31	35	35	39	39	35	35	35	40	40	40
40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	39	39	39
39	39	39	39	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40
39	39	39	40	40	40	39	39	39	33	33	34	34	40	40	25	40	40	40	40	39	35	35	35
31	26	26	26	26	35	40	33	33	33	35	35	38	40	40	40	40	40	39	35	35	35	35	35
39	40	40	39	39	39	39	39	34	34	34	33	39	24	19	12	14	19	28	25	25	24	24	24
28	28	31	32	32	33	32	33	19	19	19	19	19	29	32	33	32	33	33	25	25	25	25	25
28	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39	39
24	21	22	20	23	26	26	26	25	25	25	26	25	25	27	26	26	26	26	26	25	25	25	25
25	28	26	25	15	18	18	21	22	25	26	26	25	25	25	28	19	19	24	24	24	27	27	27
21	21	27	26	24	24	24	24	24	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
12	12	0	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	0	7	7	7	7	7
7	7	7	7	7	7	7	7	7	0	7	7	7	7	7	7	7	7	7	7	7	7	7	7
0	7	7	7	7	7	7	7	7	16														

```
>GPUPS7E01DGSP1 length=195 xy=1304_1495 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCTTTTATACGT
CCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGANCGGCATGCTTCACTTCGTGGG
GCTCAGGGGCATTTAGTCCCAGGCGCTGAACCTTNAGGCCTGAAATTTTCATGCTCNGGGC
AAAAAAAAAAAAAAAA
```

```
>GPUPS7E01DGSP1 length=195 xy=1304_1495 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 30 30 30 39 39 37
38 38 37 37 37 37 37 27 27 27 27 37 37 37 27 27 27 37 37 30 33 33
32 33 33 33 28 28 18 18 18 21 25 26 35 33 33 33 27 28 30 37 37 37
34 34 21 21 21 25 32 29 32 30 30 30 30 30 37 37 37 37 37 37 37 37
37 37 37 35 35 35 35 25 25 0 20 20 25 32 32 31 27 24 23 23 30
30 23 23 23 27 29 29 17 17 17 17 25 25 24 22 13 13 13 13 22 22 21
21 21 21 22 22 23 23 23 24 23 23 25 28 33 33 33 33 32 32 32 32
0 32 29 32 32 31 27 16 18 18 18 16 16 32 33 32 32 30 30 28 0
23 23 23 30 14 14 14 14 14 12 17 17 16 16 16 16 16 16 16 16 16 16
```

```
>GPUPS7E01BTIQD length=232 xy=0629_0643 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCTTTTATACG
TCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTG
GGGCTCGAGGGGCATTTACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTC
AGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAA
```

```
>GPUPS7E01BTIQD length=232 xy=0629_0643 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40
40 40 40 40 40 40 23 23 23 23 23 37 29 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 28 28 28 28 35 35 35 35 33 20 21 21 21
30 32 27 27 27 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
```

```

37 35 35 35 35 37 37 37 37 37 37 37 37 27 27 27 37 37 37 35 35 33
33 33 33 33 32 18 18 16 16 19 12 12 12 12 12 12 12 12 12 12 12
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 0 10 12 12 12
12 12 10 10 13 10 10 10 10 17 12 12

```

```

>GPUPS7E01BT560 length=251 xy=0636_2378 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATCTGGAATACACCAAACCACAACCTTTTTTATACGTCCCTCTCCAAACGAG
AGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATT
TACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGACAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANAAAAAAAAAN

```

```

>GPUPS7E01BT560 length=251 xy=0636_2378 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 39 39 39 39 39 39
37 25 25 23 35 33 29 32 32 32 31 31 16 18 18 18 18 20 21 33 33 33
32 23 25 25 35 37 39 37 37 25 25 25 35 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35 37 37 37
37 37 33 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 36 36 36
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 21 21 21
35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 0
12 12 12 12 12 12 12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 10
10 10 0 10 10 10 10 10 10 10 10 10 0 12 12 12 12 12 12 12 12 12
0 10 10 10 10 10 10 10 0

```

```

>GPUPS7E01CGLVD length=249 xy=0892_0551 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGGCTTGGCCACTGGAATACACCAAACCACAACCTTTTTTATACGTCCCT
CTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCT
CGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGA
CAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAA
AANAAAAAA

```

```

>GPUPS7E01CGLVD length=249 xy=0892_0551 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 39 39 39 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 38 37 38 38 38 34
34 21 21 21 21 21 23 24 35 35 35 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 30 30 30 30 37 37 37 37 37 35 35 35 33 37 37 37 37 37
37 37 37 37 30 30 30 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 33 32 30
23 23 23 27 27 14 14 14 12 12 18 12 12 12 12 12 12 12 12 12 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 12 12 12
12 12 12 12 12 0 12 12 12 12 12 12 12 12 0 12 15 15 15 15 10 10
0 10 12 15 12 10 10

```

```

>GPUPS7E01DL1L5 length=210 xy=1364_0539 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGAGAAGAGGAGGGCTTGGCCACTGGAATACACCAAACCACAACCTT
TTTATACGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCA

```

CTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATT
TCATGCTTAGGGAACAAAAAAAAAAAAAAAAA

>GPUPS7E01DL1L5 length=210 xy=1364_0539 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 39 37 32 35 35 35 32 32 21 21 21 21 21 21 25 26 35
35 35 33 28 30 37 37 37 37 37 37 21 21 21 31 37 37 37 37 37 37 37
37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 33 33 33 33 37
37 37 37 37 32 32 32 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 21
21 21 37 37 37 37 37 37 37 37 30 30 30 28 26 24 24 24 25 11 11 11
11 11 11 11 11 11 11 11 11 11 11 11

>GPUPS7E01EELT5 length=172 xy=1689_1915 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTT
ACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTCACTGCTCAGGGGACAAAAAAAAAA
ANAAAAAAAAAAAAAAAAANAAAAAAAAANAAAAAAAAANAAAAAACT

>GPUPS7E01EELT5 length=172 xy=1689_1915 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 39 37 35 35 40 39 39 38 37 37 35 34 36 27 27
27 27 37 37 37 37 37 28 28 28 28 37 37 37 37 37 35 35 35 37 37 37
37 37 37 37 37 37 37 37 37 37 31 31 31 31 31 31 37 37 37 37 35 35
35 35 30 30 30 21 21 18 27 24 21 22 25 27 25 24 22 11 11 11 11 20
25 11 11 11 11 11 11 11 11 11 11 0 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 0 12 12 12 12 12 12 12 12 12 12 0 12
12 12 12 12 12 12 12 12 0 12 12 12 12 12 12 12 14 14

>GPUPS7E01C875V length=155 xy=1218_0513 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGCATGCTTCACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGC
TGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>GPUPS7E01C875V length=155 xy=1218_0513 region=1
run=R_2010_10_21_16_08_57_
36 36 36 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 31 31 31 31 40 40 40 37 36 28 27 27 27 37 39 37
34 34 34 31 31 33 39 39 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 33 30
28 25 11 11 11 12 18 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
11 10 10 10 10 10 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 12 12 12 10 10 0 10 10 10 10 10 10 10
10

>GPUPS7E01AGE78 length=146 xy=0069_3718 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGAGAACATGCATGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGG
GCTCGAGGGGCATTTACGTCCCAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAG
GGCAAAAAAAAAAAAAAAAAAAAAAA

```
>GPUPS7E01AGE78 length=146 xy=0069_3718 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 31 31 31 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 39 39 39 37 32 32 32 33 25 25 23 31 15 13 13 13 28 25 24 22 22
11 11 11 11 25 25 16 16 11 14 14 19 14 23 23 23 28 30 32 31 31 33
35 35 32 27 27 27 27 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 35 37 37 37 37 25 21 21 21 23 16 21 21 21 21 21 21 14 14
14 14 33 33 30 28 28 22 22 22 19 18 16 16
```

```
>GPUPS7E01B91TY length=224 xy=0817_1844 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGTCCCTCTCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGC
TTCACCTTCGGTGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGA
AATTTTCATGCTCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01B91TY length=224 xy=0817_1844 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 39 39 39 40 40 40 40 40 39
39 39 40 40 40 40 40 40 39 39 39 40 40 40 40 40 39 21 21 21 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 28 26 26 26 33 33 33 32 25 18 18 18 18 32 33 28 28
28 32 33 33 37 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 26 26 26 37 37 37 37 37 37 37 37 33 32
30 21 18 18 23 25 11 11 11 11 11 11 11 11 11 11 11 11 11 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12
12 12 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 12 12 0 10 10 10 10 10 10 10 0 10 10 12 12 10 10 0
10 10 10 10
```

```
>GPUPS7E01A3X1U length=180 xy=0337_3472 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGC
TCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAATTTTCATGCTCAGGGA
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E01A3X1U length=180 xy=0337_3472 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 30 30 30 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 39 37 37 37 21 21 21 21 37 37 37 35 32 26 26
24 26 35 35 25 25 25 25 32 34 32 34 34 34 35 35 35 37 37 37 37 37
37 28 28 28 28 37 37 37 33 33 32 32 32 33 30 21 21 21 21 21 27 34
37 37 35 33 33 33 28 28 28 33 33 12 12 12 12 12 12 12 12 11 12 12
12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10
```

```
>GPUPS7E01DO5CX length=231 xy=1399_1999 region=1
run=R_2010_10_21_16_08_57_
ACGACACGTATGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCTTTTTATACG
TCCCTCTCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTG
GGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTC
AGGGAACAAAAAAAAAAAAAAAAAAAAAAAAANAAAAAAAAATAAAAAATAAA
```

```
>GPUPS7E01DO5CX length=231 xy=1399_1999 region=1
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 38
32 35 35 35 32 32 21 18 18 16 16 16 19 32 32 30 28 23 25 25 37 37
37 37 37 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 32 32 32 32 37 37 37 37 37 30 30 32 30
37 35 26 25 25 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 35 37 37 37 37 37 30 21 21 21 34 34 28 33 32 33
33 33 33 33 19 16 15 15 17 17 28 12 11 11 11 11 11 11 11 11 10 10
10 10 10 10 10 10 10 10 10 10 10 10 0 10 12 12 12 12 12 12 10 10
13 10 10 12 12 12 10 17 10 10 10
```

Lbr_NSL_PRO

```
>GPUPS7E02HR03W length=449 xy=3072_2314 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGCCGCTTTTGGTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGA
CGAGATCCAAGCTGCCAGTAGAATTCAGAATTGTCCATAGGATAGCAAACCTCATCGGCG
GGTTTTACCCAACGGTGGGCCGCATTCGGTTCGAATTCTTCTCTGCGGGATTCTTTGTAA
TTGCACAAGGTGAAATTTTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCG
ACACGCGCACTACAATGTCAGTGAGAACAAGAAAAACGACTTTTGTGCAACCTACTTGAT
CAAAAGAGTGGGGAAACCCCGGAATCACATAGACCCACTTGGGACCGAGGATTGCAATTA
TTGGTCGCGCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAACGTGCGCGATTACGTC
CCTGCCATTTGTAACAAAAA
```

```
>GPUPS7E02HR03W length=449 xy=3072_2314 region=2
run=R_2010_10_21_16_08_57_
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 33 21 21 21 21 36 36
37 39 39 39 39 39 40 40 40 40 40 40 25 25 25 38 23 23 20 35 35 31
31 39 39 33 33 33 34 34 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 30 30 30 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 33 33 33 32 33 23 23 23 35 35
36 38 37 38 23 23 23 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
34 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 35 35 26 22 22 22 22 22 38 38 19 33 25 25 25 25 39 25
39 35 35 35 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
31 17 17 17 17 31 31 31 26 26 26 26 21 21 27 25 31 31 39 39 39 39
39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
31 31 39 38 35 35 35 35 30 30 32 32 31 32 18 18 18 31 39 35 28 28
28 39 39 39 39 39 24 24 24 28 39 39 39 39 24 24 24 28 39 39 39 39
39 39 39 39 32 24 24 24 31 39 39 39 39 39 39 39 39 39 39 39 39
23 21 24 24 25 26 25 25 25 25 25 25 22 12 12 12 17 12 12 12 12 12
12 12 12 12 12 12 12 12 12
```

```
>GPUPS7E02GGL7T length=574 xy=2532_2023 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATCTGGGGCGTCAGAGGTGAAATTCTTAGACCCGCACCAAGACGAACTACAG
CGAAGGCATTCCTCAAGGATACCTTCCTCAATCAAGAACCAAAGTGTGGAGATCGAAGAT
GATTAGAGACCATTGTAGTCCACACTGCAAACGATGACACCCATGAATTGGGGATCTTAT
GGGCCGGCTGCGGCAGGGTTTACCCGTGTGTCAGCACCCGCGCCGNTTTTACCAACTTAC
GTATCTTTTCTATTTCGGCCTTTTACCGGCCACCCACGGGAATATCCTCAGCACGTTTCT
GTTTTTTCACGCGAAAGCTTTGAGGTTACAGTCTCATGGCTTCCGCGTTGCTCTTCTTTT
CTCCCTCTTGTGGGCTGGATCAGGTGCGCGCGGCGTAGTGCGGCGTAGTGCTGTGTGTGT
GTCTCTCTTGTGCAGGGACTTGTAGGTTAGTAGCAAACCTCATTCCCCTNTCGTAGCGG
AAATTA
```

```
>GPUPS7E02GGL7T length=574 xy=2532_2023 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 36 36 36 36 40 40 39 39 37
39 37 37 37 39 39 31 31 31 33 33 37 37 38 39 39 39 37 39 39 39 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 30 30 30 30 37 37 34 30 30 30 30 30 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 36 36 32
```

31	31	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37	35	35	35
37	37	35	37	37	30	27	27	27	37	35	35	34	37	37	37	37	30	28	33	33	29	
33	33	33	28	18	18	18	18	25	33	33	32	32	32	32	24	28	32	32	37	37	37	
37	37	37	37	37	37	37	37	37	37	37	35	35	34	37	35	35	35	35	35	30	30	
30	25	27	27	37	37	37	37	32	35	37	37	37	37	37	37	37	37	37	37	37	34	
35	35	35	35	37	0	28	28	30	30	37	37	37	37	37	37	31	31	30	37	37	37	
35	33	33	24	24	24	24	27	18	24	19	17	18	24	27	17	19	15	15	15	13	13	
18	23	30	27	27	26	22	13	15	15	30	24	15	15	13	13	13	13	18	28	27	27	
27	21	24	24	30	30	30	27	17	17	17	17	21	16	22	12	12	11	11	11	12	22	
25	27	27	21	21	23	23	26	30	32	28	28	28	28	32	33	33	33	35	35	35	35	
30	28	28	28	32	32	33	31	33	31	23	23	23	23	27	27	30	29	29	32	32	19	
15	15	15	23	18	20	18	18	31	18	23	23	23	23	17	29	29	30	29	29	29	29	
35	35	35	35	35	35	35	35	35	35	35	29	28	28	23	23	23	29	33	33	33	31	
31	29	29	29	23	23	23	27	29	27	27	27	29	23	23	23	28	28	28	28	29	29	
29	24	24	24	25	22	22	24	24	23	23	22	22	22	19	22	17	11	11	11	11	11	
15	15	15	11	11	11	13	13	11	11	11	13	13	19	13	12	12	11	15	18	22	22	
17	17	17	12	12	12	12	15	0	23	25	25	23	25	22	22	22	22	19	15	15	15	
15	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	0	7	7	7	
7	7	7	7	7	7	7	7	7	7	7	0	7	7	7	7	7	7	7	7	7	7	
7	7	7	7	7	0	7	7	7	7	7	7	7	7	7	7	7	7	7	0	7	7	
7	7	7	7	7	7	7	7	7	7	0	7	7	7	7	7	7	7	7	7	25	7	7
7	7																					

```
>GPUPS7E02FFUJJ length=338 xy=2113_2765 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATCGAACTACAGCGAAGGCATTCTTCAAGGATACCTTCCTCAATCAAGAAC
CAAAGTGTGGAGATCGAAGATGATTAGAGACCATTGTAGTCCACACTGCAAACGATGACA
CCCATGAATTGGGGATCTTATGGGCCGGCCTGCGGCAGGGTTTACCCTGTGTGTCAGCACCG
CGCCCCGTTTTACCAACTTACGTATCTTTTCTATTTCGGCCTTTACCGGCCACCCACCGGG
AATATCCTCAGCACGTTTTTCTGTTTTTCTACGCGAAGACTTTGTAGGTACTAGCGTCAGG
GGGGAGGGTACTTCGCAAGAAGTGAAACAAAAAAAAA
```

```
>GPUPS7E02FFUJJ length=338 xy=2113_2765 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 30 30 30 30 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 26 26 26 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 31
31 31 36 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 28 28
28 28 32 32 34 37 37 37 37 37 37 37 37 31 31 31 31 31 37 37 37 37 37
37 37 37 37 30 37 37 37 37 30 23 21 21 21 18 27 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 31 31 28 28 37 37 37 37 37 37 34 32
29 31 32 32 32 32 27 24 16 16 16 16 20 18 31 32 33 32 32 33 35 37
37 37 37 37 35 34 30 28 28 28 33 28 26 26 16 16 16 16 16 26 23 23
24 27 30 28 23 23 24 24 24 31 33 33 33 26 26 26 22 25 15 32 20 22
20 17 17 27 16 19 19 18 13 13 13 12 12 20 18 11 11 11 11 11 12 13 22 12
12 12 12 18 12 12 12 12 12 12 12 11 11 11 11 11 11 11 24 9 9 9 24
24 25 21 17 13 13 13 17 21 22 16 16 16 18 16 16 16 16 22 11 11 11
11 11 11 11 11 11
```

```
>GPUPS7E02JJELS length=450 xy=3794_1822 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGCCGCTTTTGGTGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGA
CGAGATCCAAGCTGCCAGTAGAATTGAGAATTGTCCATAGGATAGCAAACCTCATCGGCG
GGTTTTACCCAACGGTGGGCCGCATTGGTTCGAATTCTTCTCTGCGGGATTCCTTTGTA
TTGCACAAGGTGAAATTTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCG
ACACGCGCACTACAATGTCAGTGAGAACAAGAAAAACGACTTTTGTGCAACCTACTTGAT
CAAAAGAGTGGGGAAACCCCGGAATCACATAGACCCACTTGGGACCGAGGATTGCAATTA
TTGGTCGCGCAACGAGGAATGTCTCGTAGGCGCAGCTCATCAAACGTGTCCGATTACGTC
CCTGCCATTTGTAACAAAAA
```

```
>GPUPS7E02JJELS length=450 xy=3794_1822 region=2
run=R_2010_10_21_16_08_57_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 34 34 34 34 40 40
40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 31 31 31 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39
40 40 40 40 40 40 40 40 39 39 39 34 34 34 34 40 39 39 39 40 40
40 40 40 40 29 29 29 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 39 39 39
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 31 31 31 31 31 31
31 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 35 35 22 22 22 22 22 39 40 26 40 34 34 34 34 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 34 34 34 34 40 22 36
35 22 22 21 21 15 15 14 14 14 14 14 16 16 22 22 32 33 33 36 38 40
37 40 37 27 27 27 38 31 35 35 27 27 27 38 31 31 24 38 38 38 37 31
31 31 40 31 31 25 25 23 33 33 35 35 35 35 36 36 39 33 31 31 29 33
33 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39 39
39 39 39 39 39 31 31 31 39 39 39 39 39 39 39 39 39 39 39 39 25 25
23 24 27 27 27 26 26 26 27 23 23 20 24 11 15 15 20 11 11 11 11 11
11 11 11 11 11 11 11 11 11 11
```

```
>GPUPS7E02GC8AZ length=451 xy=2493_3769 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGTTAAAGGGTTCGTAGTTGAACTGTGGGCTGTGCAGGTTTGTCCCTGGT
CGTCCCGTCCATGTCGGATTTGGTGACCCAGGCCCTTGCAGCCCGTGAACATTCAAAGAA
ACAAGAAACACGGGAGTGGTTCCTTTCTGATTTACGCATGTCATGCATGCCAGGGGGCG
TCCGTGATTTTTTACTGTGACTAAAGAAGCGTGACTAAAGCAGTCATTCGACTTGAATTA
GAAAGCATGGGATAACAAAGGAGCAGCCTCTAGGCTACCGTTTCGGCTTTTTGTTGGTTT
TAAAGGTCTATTGGAGATTATGGAGCTGTGCGACTCTTGCTGCGTGTGGAAGAGGGGGAG
GGCGGGGGAGGAGGGAGTTGGGGCGTAGGTGCGGTACAAAAAAAACGAAAAAAAAAAAA
NAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E02GC8AZ length=451 xy=2493_3769 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 33 33 31 31 35 37 37 37 35 35
36 37 39 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40
40 40 40 40 39 39 39 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 35 35 36 37 37 37 37 37 37 37 37 37 31 31 31 37 37 37 37
37 37 37 37 35 35 35 37 37 37 37 37 37 21 21 21 28 35 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 21 21 21
21 21 30 18 28 32 32 32 32 32 17 17 17 17 17 17 25 25 14 32 33
```



```
>GPUPS7E02HKAKA length=241 xy=2984_1768 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39
39 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 35 35 35 37 37 37 37 37 37 37
37 31 31 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
32 32 32 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 31 31 31 21 24 22 22 22 22 37 12 12 12 12 12 12
12 12 12 1212 0 12 12 12 12 12 12 12 12 12 12 12 12 0 12 12 12
12 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 0
```

```
>GPUPS7E02GP34K length=223 xy=2640_2770 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATCTTGGCTTCTTTCGTTGAAGAACGCAGTAAAGTGCATAAGTGGTA
TCAATTGCAGAATCATTCAATTACCGAATCTTTGAACGCAAACGGCGCATGGGAGAAAGCT
CTATTGTGTCATCCCCGTGCATGCCATATTCTCAGTGTGCAACAAAAAACAAAAAAAN
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN
```

```
>GPUPS7E02GP34K length=223 xy=2640_2770 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39
39 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
28 30 28 28 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
32 32 32 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
35 33 32 32 29 23 15 13 14 11 11 11 11 11 11 23 11 11 11 11 11
11 11 11 0 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
10 10 10 10 10 10 10 10 10 10 12 12 12 12 12 12 10 0 10 10 10
10 10 0
```

```
>GPUPS7E02IJPKY length=468 xy=3387_3504 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATAGCGGAGCAGGCGATGCCGAATGAACCGCTTAGGGATATGTGGCGTCAA
AACTTATGGGCTCATTCTCGATACGTGAAAAGGGTGTGGTGGATATGGACAGTTGGACG
GTGGCCATGGAAGTCGGCATCCGCTAAGGAGTGTGTAACAACCTACCAACCGAATCCACC
GGCCCCGAAAATGGATGACGCTTAAGCCCAATTAGTGATTGCCATTATTCCCTTTGGTAA
GGCGGAAACATACTCTTGGAGAAAGTGTAGCCCTTCCGATGTAGATGTGGCCTGGAGGT
TAGGACGAAGCTTATGGCGTGAGCCTAAGATGGACCGCCTCTAGTGCAGATCTTGGTTG
GCGTAGCAAAGATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAN
```

```
>GPUPS7E02IJPKY length=468 xy=3387_3504 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39
39 39 40 40 40 40 40 40 37 37 37 37 37 37 35 35 35 35 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 31
31 31 31 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 31 31 31 37 37 37 37
```



```
>GPUPS7E02GRRJQ length=559 xy=2659_1956 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGCTAGTAGAAGTATTGTGCGGTAAAGCAAATGATTAGAGGCATTGGTGT
TCTTAGAATATCGACCTATTCTCAAACCTTTAAATGTGCAAACAAACCGTGCCTTAGCCAA
CTGCAAGGCCGGAAGAAAAGAATTAGAGGCACCAAGTGGGCCTTCTTCGGTAAGCGGAG
CAGGCGATGCGGAATGAACCGCTTAGGGATATGTGGCGTCAAACCTTATGGGCTCATTCT
CGATACGTGAAAAGGGTGTGGTGGATATGGACAGTTGGACGGTGGCCATGGAAGTCGGC
ATCCGCTAAGGAGTGTGTAACAACCTACCAACCGAATCCACCGGCCCGAAATGGATGAC
GCTTAAGCCCAATTAGTGATTGCCCATTTCTTTGGTAAGGGCGGAAACATACTCTTG
GAGAAAGTGTAGCCCTCCGATGTAGATGTGGCCTGGAGGTTAGGACGAAGCTTATGGCG
TGAGCCTAAGATGGACCGCCTCTAGTGCAGATCTTGGTTGGCGTAGCAAAGATACTAAA
CAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E02GRRJQ length=559 xy=2659_1956 region=2
run=R_2010_10_21_16_08_57_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 35 35 35 40 40 40
38 38 38 40 40 40 40 40 40 40 40 40 40 40 38 38 38 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 38
38 38 39 39 39 40 40 40 40 40 26 26 26 34 39 39 39 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 39 39 40 22 22 22 22 38 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 40 35 35 33 38 38 40 40 40 40 40 40 40 40
33 33 33 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 38 38
38 38 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40
34 34 34 34 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 31 31 31 31 39 39 39 40 40 40 40 39 40 40 38
38 38 40 40 40 33 33 33 40 40 40 40 40 40 40 40 40 40 40 38 38 38
40 40 40 40 40 40 40 40 38 33 35 35 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 39 35 33 33 33 39 35 40 38 40 40 39 39 40
40 40 39 38 35 35 38 35 35 35 32 32 32 32 14 14 14 14 17 14 14 14
22 26 29 31 33 33 33 36 36 40 40 40 40 40 40 39 39 39 40 40 40 40
40 40 40 40 40 40 40 40 37 37 36 36 36 36 32 26 26 27 26 25 23 23
19 19 12 16 24 11 11 11 24 11 12 14 14 14 15 17 32 30 30 28 28 33
33 28 30 25 22 14 14 13 24 27 23 23 23 22 13 13 13 17 17 17 17 27
23 23 23 24 27 25 27 30 27 22 22 22 25 25 25 28 30 27 23 23 19 19
18 25 26 30 19 13 14 14 30 30 30 29 29 26 29 30 30 29 30 28 29 27
29 29 29 19 19 19 22 29 26 28 28 30 28 23 23 23 23 25 20 22 25 28
22 18 19 19 19 18 19 25 22 22 26 26 19 20 20 20 28 23 25 25 22 25
25 25 25 25 16 25 22 25 23 28 28 27 27 12 12 12 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12
```

```
>GPUPS7E02HOFE3 length=467 xy=3031_2173 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGCGAGCTCAGTGGGAACAGAAATCTCACGTAGAACACAAGCGTAAAAGC
TTGCTTGATTAACGATTTCCAGTACGAATCGAGACTGCGAAAGCAAGGCCAGCGATCCTT
TGCGAAGAATAGGAAATATGAACAAATCATACCAGAGGTGTCAGAATAATTACCACAGGG
ATAACTGGCTTGTGGCGGCCAAGCGTTCATAGCGACGTCGCTTTTTTGATCCTTCGATGTC
GGCTCTTCCCTAACCTAGCGCCGAGAAGACGCTAAGGGTTGGATTGTTACCCACTGACA
GGGAACGTGAGCTGGGTTTAGACCGTCGTGAGACAGGTTGTTTACCCTACTTAGCTAG
GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E02HOFE3 length=467 xy=3031_2173 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
31 31 31 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40
40 39 37 37 37 37 37 38 34 27 21 21 21 21 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 35 35 35 37 37 37 37 37 37 37 36 36 36 37 37 37 37
37 37 37 37 37 37 37 37 36 36 36 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 35 31 31 32 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 31 31 31 37 37 37
37 30 30 30 28 28 37 37 37 37 37 37 37 37 37 37 37 34 30 26 26 26
37 37 37 35 35 34 34 34 37 37 37 37 38 38 37 37 35 35 35 35 35 35
34 34 38 35 35 37 38 38 35 35 35 35 35 38 38 38 30 30 37 35 35 38
38 38 37 37 34 30 28 33 32 32 29 29 27 23 19 19 19 19 14 13 12 18
24 24 24 24 23 11 11 11 11 13 13 18 14 20 16 16 16 22 26 30 32 31
28 25 25 24 32 32 32 32 29 29 29 19 17 19 19 24 14 14 13 12 17 14
17 17 30 30 25 18 18 19 23 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 0 7 7 7 7 7 7 7 7 7 0 7 7 7 7 7 7 7
7 7 7 7 7 7 7 7 7 7 7 7 7 7 0 7 7 7 7 7 7
7 7 7 7 0 7 7 7 7 7 7 7 7 7 0 7 7 7 7 7 7
7 7 0 7 7 7 7 7 7 7 7 0 7 7 7 7 7 7 7 7 0
7 7 7 7 7
```

```
>GPUPS7E02IUMQ6 length=176 xy=3512_1056 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTC
ACTTCGGTGGGGCTCGAGGGGCATTTACGTCCCAGGGCGCTGAACCTTGAGGCCTGAAAT
TTCATGCTCAGGGACAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAACACAC
```

```
>GPUPS7E02IUMQ6 length=176 xy=3512_1056 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 28 28 30 30 37 37 37 37 37 32 35 32 35 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37
30 33 32 31 28 27 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 0 10 10 10 10 10 10 10 10 10 10 13 10 10 10 12 13 24 23 23
```

```
>GPUPS7E02FPUSB length=185 xy=2227_2697 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGC
TCGAGGGGCATTTACGTCCCAGGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGG
ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAA
TAAAG
```

```
>GPUPS7E02FPUSB length=185 xy=2227_2697 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40 40
40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 39 35 35 35 25 25 25 25 37 37 37 35 35 28 28
28 28 37 37 26 26 26 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
```

```

37 37 37 37 37 37 28 21 21 21 26 26 12 12 12 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 0 10 10 10 10 10 10 10 10 10 10 10 10 10 13 10 10 12
12 12 10 10 13 10 10 10 13

```

```

>GPUPS7E02GPH29 length=216 xy=2633_2883 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGGAGAGGGCTTGGCCACTGGAATACACCAAACCACAACCTTTTTATA
CGTCCCTCTCCAAACGAGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGG
TGGGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGC
TCAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```

>GPUPS7E02GPH29 length=216 xy=2633_2883 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 39 39 39 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39
37 35 32 35 35 35 32 32 18 18 18 18 16 23 21 32 32 32 32 23 23 25
35 37 37 37 37 25 25 25 28 35 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 36 36 35 30 30 30 30 37 37 37 37 37 28 28
28 31 37 37 25 25 25 37 37 37 37 35 35 35 37 37 37 37 37 37 37
37 28 28 28 28 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 31 31 31 36 12 12 12 12 12 12 12 12 12
12 12 12 12 12 15 15 15 15 15 15 10 10 11 10 10 10

```

```

>GPUPS7E02GGFXE length=187 xy=2530_2064 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGC
TCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGG
ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
NAAAAAA

```

```

>GPUPS7E02GGFXE length=187 xy=2530_2064 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 37 35 32 32 32 37 37 37 37 36 31 31
31 31 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 25 25 25 30 14 12 12 12 12 12 12 12 10 10
10 10 10 10 10 10 10 10 10 10 12 12 16 21 21 21 21 21 21 21 21
21 15 15 10 10 10 10 10 10 10 10 10 10 10 10 10 10 0 10 10 10
10 10 10 10 0 10 10 10 10 10

```

```

>GPUPS7E02GNWVT length=234 xy=2615_2471 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATGAGAGAACATGCATGGGCTGGCATGAGCGGCATAGCTTTCCTTCGGTG
GGGCTCGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCT
CAGGGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```
>GPUPS7E02GNWVT length=234 xy=2615_2471 region=2
run=R_2010_10_21_16_08_57_
30 30 34 37 37 37 32 32 32 32 35 35 32 32 32 32 28 25 25 24 27 30
31 27 27 26 26 26 19 19 19 19 30 32 31 27 19 19 19 21 24 24 27 28
28 25 19 13 13 13 18 19 17 16 22 23 20 20 20 13 13 13 13 25 29 23
24 19 19 19 19 21 33 33 28 28 28 32 21 21 23 34 34 37 37 34 34 32
37 37 37 37 37 35 35 35 35 37 37 37 37 35 33 32 32 32 33 28 25 23
16 16 16 16 24 32 30 30 27 25 24 28 18 18 18 20 22 11 11 11 11 11
11 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12
12 10 10 0 10 10 12 12 12 12 12 12 12 12 12 12 12 12 12 0 12 12
12 12 12 12 12 12 12 12 12 10 10 0 10 10 10 10 10 10 10 10 10
10 0 10 10 10 10 10 10 10 10 10 10 10 0 10 10 10 10 10 10 10
10 0 10 10 10 10 10 10 10 10 13 10 10 10 10
```

```
>GPUPS7E02F0JN1 length=216 xy=2349_1791 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATAGAACATGCATGGGCTGGCATGAGCGGCATGCTTCACTTCGGTGGGGCT
CGAGGGGCATTTACGTCCCGAGGCGCTGAACCTTGAGGCCTGAAATTTTCATGCTCAGGGA
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

```
>GPUPS7E02F0JN1 length=216 xy=2349_1791 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 39 39 37 37 28 28 28 28 37 35 33 33 28 20 20 20
20 35 32 21 21 21 32 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 25 25 25 35 35 35 37 37
37 37 37 37 37 37 30 30 30 36 37 12 12 12 12 12 12 12 12 12 12
12 12 12 12 12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 10 10
12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10 0 12 12 12 12
12 12 12 12 12 12 0 10 10 10 10 10 10 10 10 10 0 10 10 10 10
10 10 10 0 10 12 12 12 12 10 10 0 10 10 10 10 10 0
```

La_PH8_MET

```
>GWO0T3C01COX8Y length=159 xy=0987_0720 region=1
run=R_2011_01_26_16_16_41_
GGCCGTGTAGTGGACTGCCATGGCGTTGACGGGAGCGGGGATTAGGGTTCGATTCCGGA
GAGGGAGCCTGAGAAATAGCTACCACCTTCTACGGAGGGCAGCAGGCGCGCAAATTGCCCC
AAATTGGTTCCAAAAAAAACCAAAAAAAAACCGGGAAA
```

```
>GWO0T3C01COX8Y length=159 xy=0987_0720 region=1
run=R_2011_01_26_16_16_41_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 39 39 39 39 39 40 30 30 30 28 28 39 39 39
39 39 39 39 39 39 39 39 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 26 26 26 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 35 33 32 32 32 32 27 19
15 15 13 13 13 18 14 14 14 14 16 16 16 16 16 16 16 20 20 12 12 11
11 11 11 11 11 11 11 11 11 11 10 10 10 10 10 10 10 10 10 10 11 11 11
11 11 10 10 10
```

```
>GWO0T3C01DH9K0 length=341 xy=1321_0370 region=1
run=R_2011_01_26_16_16_41_
CAAGACGTGGAGCGTGC GGTTTAAATTTGACCCAACACGGGGAAC TTTACCAGATCCGGAC
AGGATGAGGATTGACAGATTGAGTGTCTTTCTCGATTCCCTGAATGGTGGTGCATGGCC
GCTTTTGGTTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAACGGACGAGATCCAAGCTG
CCCAGTAGAATTCAGAATTATCTGCTTCCATCCGTGACAAGACATGAGGAAACCGAAGTA
ACCAACGGAAGAAAACACGTGTGAGGGTGAAAAGAGGAAGACACGAGCAACATGATCACA
GCACAGCACCAAACGCGAAGGAAAAAGAAAAA AAAAAAAAAA
```

```
>GWO0T3C01DH9K0 length=341 xy=1321_0370 region=1
run=R_2011_01_26_16_16_41_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39
40 40 39 39 39 40 40 39 39 39 40 40 40 40 40 40 31 31 31 31 40 40 40
39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
39 39 39 40 40 40 40 40 40 40 38 38 38 40 40 40 40 40 40 40 40
40 40 40 40 36 36 34 38 40 40 40 40 34 34 34 34 40 40 40 40 40 40
40 40 40 40 40 40 40 40 39 39 39 40 35 35 35 40 40 40 40 40 40 40
40 40 40 40 37 37 34 34 34 40 40 40 40 40 40 40 40 40 30 30 30 30
30 40 40 40 37 39 39 40 40 40 40 40 40 40 40 40 40 38 38 34 34 34
34 40 40 40 40 40 40 40 38 38 34 34 34 40 40 40 40 38 34 34 34
40 40 36 36 36 40 40 23 23 18 20 20 19 19 36 27 27 34 39 36 36 34
33 17 16 16 22 33 33 27 24 14 14 14 14 22 17 33 36 37 28 23 23 28
24 24 31 31 16 14 14 14 14 22 20 34 34 31 28 26 26 32 36 40 40 36
36 29 29 29 28 26 36 33 33 33 36 33 32 32 38 36 32 31 31 35 32 27
26 26 17 17 17 19 25 28 26 24 27 26 26 17 19 20 20 20 40 12 12 12
12 12 12 12 12 12 12 12 12 12 23
```

```
>GWO0T3C01BPZKL length=390 xy=0588_3779 region=1
run=R_2011_01_26_16_16_41_
CTTTACCAGATCCGGACAGGATGAGGATTGACAGATTGAGTGTCTTTCTCGATTCCCTG
AATGGTGGTGCATGGCCGCTTTTGGTTCGGTGGAGTGATTTGTTTGGTTGATTCCGTCAAC
GGACGAGATCCAAGCTGCCAGTAGAATTCAGAATTGCCCATAGAATAGCAAACATCATCG
GCGGGTTTACCCAACGGTGGGCCGCATTCGGTTCGAATCTTCTCTGCGGGAGTAATAAT
ATATATATATAAGGACACAAGCGCACAGGCCGAAACAGACAACACACGCCGGGAAANAAAA
CCACACACACATACATAACA AAAAAAAAAAGGAAACAAAAACAAA AAAAAAAAAACA
AACGCTACGCATGTAACGTCAAAAACCGAA
```

```
>GWO0T3C01BPZKI length=390 xy=0588_3779 region=1
run=R_2011_01_26_16_16_41_
37 35 35 35 35 35 37 37 37 37 37 37 37 37 34 35 36 37 37 37 37 37 37
37 37 38 38 39 39 39 39 39 39 39 39 40 39 39 37 35 35 35 35 35 35 32
21 21 21 21 25 29 35 33 33 33 33 28 30 37 37 28 28 28 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 31 31 31 31 37 37 37 37 37
37 35 25 25 27 27 27 37 37 37 37 37 30 29 31 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 35 35 37 37 34 34 34 35 35 37 37 37 37 37
37 37 37 37 37 37 37 37 37 35 32 31 31 31 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 35 35 35 35 33 33 31 30 16 16 15 18 22 27
27 22 22 20 20 12 11 11 11 11 11 14 13 13 11 11 11 12 12 11 20 20
28 11 13 13 19 19 22 27 30 29 32 32 32 33 33 33 35 35 35 35 32 21
21 21 32 32 32 35 35 32 30 30 30 32 26 32 32 32 35 35 35 32 28 28
27 30 32 31 31 27 27 27 19 19 19 19 19 19 23 31 32 32 30 24 23 23
24 27 28 29 29 29 27 15 15 15 19 15 19 19 28 26 21 18 17 18 18 21
22 16 16 16 16 16 11 11 11 0 11 11 11 11 12 12 11 16 24 24 27 27
27 27 28 28 28 24 21 22 20 22 20 22 16 16 16 11 11 11 11 11 11 11
11 20 20 12 12 12 24 13 14 14 14 15 27 10 10 10 10 19 24 12 12 12
12 12 12 12 12 12 31 7 7 7 26 25 27 29 29 26 23 18 18 18 18 29
33 33 33 31 31 31 16 16 16 16 16 23 23 18 9 9
```

```
>GWO0T3C01BKANS length=375 xy=0524_0358 region=1
run=R_2011_01_26_16_16_41_
CTTTACCAGATCCGGACAGGATGAGGATTGACAGATTGAGTGTTCTTTCTCGATTCCCTG
AATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTGGTTGATTCCGTCAAC
GGACGAGATCCAAGCTGCCAGTAGAATTCAGAATTGCCCATAGAATAGCAAACATCATCG
GCGGGTTTACCCAACGGTGGGCCGCATTCGGTCAATTCTTCTCTGCGGGAGTAATAAT
ATATATATATAAGGACACAAGCGCACAGGCCGAAAACAGACAACACACGCCGGGAAGAAAA
CCACACACACATACATACACAAAAAAAAGGAAAAACAAAAACAAAAACTAAAAAAA
AACAAACGACTACGA
```

```
>GWO0T3C01BKANS length=375 xy=0524_0358 region=1
run=R_2011_01_26_16_16_41_
37 37 37 35 35 35 35 37 37 37 37 37 37 37 37 39 39 39 39 39 40 40
37 39 39 39 39 39 40 40 39 39 39 39 39 40 40 40 40 40 40 40 40
40 39 39 39 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 35 32 32 32 37 37 37 37
25 25 35 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 31 23 23 25 31 35 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 34 30 30 35 35
37 27 21 21 21 34 34 37 37 32 28 28 28 37 37 37 37 37 35 24 27 27
32 37 37 37 37 37 37 35 35 35 35 37 37 37 37 37 37 35 35 35 37
37 37 37 37 37 37 35 26 26 26 37 37 37 37 34 32 32 35 35 37 35 37
37 37 37 35 37 37 37 37 37 32 32 32 32 32 35 35 35 32 28 23 23
24 28 28 26 26 27 21 13 13 13 13 13 12 18 18 27 26 24 24 27 30 30
30 30 29 29 21 21 21 17 12 13 13 13 13 13 13 13 13 23 27 27 27 24
23 23 23 23 32 30 30 27 27 22 16 18 16 20 11 11 11 11 11 11 11 11
11 20 20 10 10 10 10 24 10 10 10 10 10 10 16 9 9 9 9 9 16 18
10 10 12 12 12 12 12 12 12 26 7 7 7 17 22 17 21 18 17 17 18
11
```

```
>GWO0T3C01EWSO4 length=331 xy=1896_2742 region=1
run=R_2011_01_26_16_16_41_
TGCGACAAGTGCTTTCCCATCGCAACTTCGGTTCGGTGTGTGGCGCCTTTGGAGGGGTTT
AGTGCGTCCGGTGCAGCTCCGGTTCGTCCGGCCGTAACGCCTTTTCAACTCACGGCCTC
TAGGAATGAAGGAGGTAGTTTCGGGGGAGAACGTACTGGGGCGTCAGAGGTGAAATTTGA
GATACATAGAAGGAAGCAAAGAAACGAACGCACGCACGCAGATACACAAGAATGGAAAA
CCACAAAGGAAGAACAACAGACAGAGAGAGAGATGTATCGCGTGCGTGGAGGTGGAA
AGTAAACGAAAAAAAAACGAAAAAAAAAGAAG
```

```
>GWO0T3C01EWSO4 length=331 xy=1896_2742 region=1
run=R_2011_01_26_16_16_41_
37 37 37 37 37 37 37 37 37 37 37 39 39 37 37 30 30 30 37 37 37 37
33 30 30 33 35 37 39 39 40 37 39 39 39 40 40 40 40 40 39 37 37
37 35 33 28 28 26 31 29 24 16 16 16 16 25 25 30 35 26 32 35 35 37
37 37 37 37 37 35 31 32 31 35 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 35 35 35 35 37 37 37 37 32 32 36 36 37 37 37
37 37 37 37 37 37 37 37 37 35 35 35 37 37 37 37 37 37 37 37 23 23
25 31 30 37 37 37 32 35 35 35 22 22 22 22 22 30 19 30 28 32 28 28
28 32 32 21 21 21 24 28 22 28 35 35 32 30 30 30 37 37 37 37 37
37 37 37 37 37 37 37 35 33 33 32 32 32 19 19 19 19 19 17 17 14 27
21 21 21 22 12 12 12 19 19 22 22 30 30 23 19 19 20 33 25 25 25 25
35 35 35 35 37 37 37 35 32 32 32 33 33 33 33 21 21 19 19 19 19
16 27 23 23 23 23 19 19 24 24 24 25 28 28 28 33 32 35 35 35 35
35 35 35 25 25 25 27 27 30 31 31 23 24 24 23 24 23 25 32 33 33 33
35 35 35 35 35 37 38 35 33 33 33 27 26 26 18 19 12 12 12 18 17
11 11 11 11 11 11 11 11 11 22 20 8 8 8 8 8 8 8 16 8 8 12
12
```

```
>GWO0T3C01B73UY length=468 xy=0795_1272 region=1
run=R_2011_01_26_16_16_41_
TTGGGCAACAGCAGGTCTGTGATGCTCCTCAATGTTCTGGGCGACACGCGCACTACAATG
TCAGTGAGAACAAGAAAAACGACTTTTGTGCAACCTACTTGATCAAAAGAGTGGGGAAAC
CCCAGAAATCACATAGACCCACTTGGGACCGAGGATTGCAATTATTGGTTCGCGCAACGAGG
AATGTCTCGTAGGCGCAGCTCATCAAACGTGTCCGATTACGTCCCTGCCATTTGTACACA
CCGCCCGTCTGTTTCCGATGATGGTGCAATACAGGTGATCGGACAGGCGGTGTTTTAT
CCGCCCGAAAGTTCACCGATATTTCTTCAATAGAGGAAGCAAAAGGGGGAGGCTGGAGAA
CGAGAGTGGCAGCACTGAAGAGTTTTACGTGGTGTGGGCAACAATGGAAAAAAGAAAAA
AGGAATAACAAAAAACGACAAACCAATAAAAAAACGAAAAGTAGACGG
```

```
>GWO0T3C01B73UY length=468 xy=0795_1272 region=1
run=R_2011_01_26_16_16_41_
39 39 21 21 21 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 38 38 39 38 36 35 16 16 16 16 16 22 30 26 36 25 25 25 28 40
37 40 40 40 40 40 40 40 39 37 37 37 40 40 40 40 34 34 34 34 40 37
40 40 33 33 32 32 37 37 37 32 32 32 32 37 37 39 39 40 40 40 40 40
40 40 40 40 39 39 39 40 40 40 40 39 39 39 40 33 33 32 40 40 40 40
40 40 40 40 39 40 33 33 33 39 40 39 40 39 40 40 40 40 39 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 37 33 29 29 29 40 40 40 38 38
38 39 38 38 40 33 21 21 20 22 27 38 38 38 38 38 38 38 38 37 40
38 33 21 20 20 36 36 30 30 30 18 18 18 21 19 38 38 38 37 37 37
40 39 39 39 40 40 40 40 33 33 29 29 27 40 35 35 40 40 40 40 40
40 40 40 40 40 39 39 39 40 40 27 27 27 27 35 40 40 31 31 31 30 31
35 32 32 32 21 20 18 11 14 14 14 14 15 13 23 23 24 25 25 30 38 36
39 39 40 40 40 40 40 40 40 35 27 21 21 15 15 15 20 17 22 28 23 23
24 21 24 23 16 16 16 16 21 22 11 11 11 11 11 11 11 11 11 10 10
```

```

18 22 23 26 28 26 31 31 32 32 30 24 19 19 20 24 24 34 34 30 25 25
24 25 32 32 32 31 26 23 13 13 13 13 15 22 19 14 19 19 19 15 15 11
11 11 18 15 15 15 15 15 22 11 11 11 11 11 11 11 11 15 9 9 9 9
9 8 8 11 11 7 7 11 11 11 11 13 14 17 17 14 12 12 13 8 15 15
15 15 15 15 12 12 18 12 12 15 15 15 15 14 14 27 29 7 7 7 21 25
15 16 15 18 15 14

```

```

>GWO0T3C01E07AP length=367 xy=1946_3487 region=1
run=R_2011_01_26_16_16_41_
CTTTACCAGGATCCGGACAGGATGAGGATTGACAGATTGAGTGTTCTTTCTCGATTCCCT
GAATGGTGGTGCATGGCCGCTTTTGGTCGGTGGAGTGATTTGTTGGTTGATTCCGTCAA
CGGACGAGATCCAAGCTGCCAGTAGAATTCAGAATTGCCCATAGAATAGCAAACATC
GGCGGGTTTTACCCAACGGTGGGCCGCATTCCGGTCGAATTCTTCTCTGCGGGAGTAATA
TATATATATATAAGGACACAAGCGCACAGGCGAAACAGACAACACACGCCGGGAAGAAAA
CCACACACACATACATACACAAAAAAAAAGGAAACAAAAACAAAACATAAAAAAAAAACA
AACGACT

```

```

>GWO0T3C01E07AP length=367 xy=1946_3487 region=1
run=R_2011_01_26_16_16_41_
34 23 23 23 37 37 37 30 30 30 32 35 37 37 33 33 33 36 37 37 39 39
39 39 39 39 37 39 38 38 39 39 39 39 33 28 31 31 39 37 39 39 38 38
38 38 37 35 35 37 39 36 35 35 37 37 37 37 37 34 34 34 37 37 37
37 35 34 34 34 37 37 37 37 37 35 37 37 34 28 28 28 28 32 34 37 37
37 37 37 27 27 28 28 28 37 37 37 37 37 37 37 37 37 37 37 37 37
37 34 34 30 25 25 27 34 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 34 34 32 30 30 30 30 35 35 35 35
33 35 35 35 21 21 21 37 30 31 32 35 37 37 37 37 37 37 37 37 37
27 27 27 27 27 37 33 28 27 27 21 21 21 21 20 18 18 18 23 23 22 28
30 30 23 24 24 20 21 18 30 32 32 32 33 33 33 32 32 32 32 35 35 34
30 30 30 32 34 37 37 37 37 37 30 37 37 37 37 34 34 32 32 32 32
35 35 32 30 30 30 27 27 27 29 32 19 19 19 19 19 27 27 27 24 24 23
27 29 30 32 29 29 27 23 19 19 19 19 25 29 27 27 24 24 27 26 27 27
24 22 17 17 13 13 13 21 21 14 11 11 11 11 24 24 16 23 24 19 19 19
23 23 27 27 28 31 32 33 28 25 25 25 25 30 12 12 16 16 16 16 12
12 20 20 8 8 10 27 15 15 15 15 15 30 16 16 16 16 27 27 12 12 12
12 12 12 12 12 12 19 7 7 7 11 11 14 15 18

```

```

>GWO0T3C01EIGMS length=309 xy=1733_1570 region=1
run=R_2011_01_26_16_16_41_
GGATCTTATGGGCCGGCCTGCGGCAGGGTTTACCCTGTGTCAGCACCGCGCCCGCTTTAC
CAACTTACGTATCTTTTCTATTCGGCCTTTACCGGCCACCCACGGGAATATCCTCAGCAC
GTTTTCTGTTTTTACGCGAAAGCTTTGAGGTTACAGTCTCAGGGGGAGTACGTTTCGCAA
GAGTGAAACTTAAAGAAATTGACGGAATGGCACCACAAGACGTGAGCGTGCGGTTTAATT
TGACGCAACACGGGAACCTCCACCCCTGGATCGTATGCGTATATAGATTAATAAAAAACAGA
ACAAACAAC

```

```

>GWO0T3C01EIGMS length=309 xy=1733_1570 region=1
run=R_2011_01_26_16_16_41_
37 37 37 37 37 37 37 37 37 21 21 21 37 37 37 37 38 38 38 38 37 37
37 39 39 35 31 31 31 31 31 39 39 39 39 40 40 40 40 40 40 40 40
40 40 38 38 35 35 30 30 30 35 18 21 21 18 18 20 33 33 33 33 33
35 37 37 37 37 37 37 35 35 35 35 37 37 37 37 37 37 37 37 35
35 35 37 37 37 35 32 32 29 24 19 19 19 19 16 16 16 15 19 19 30 29
32 32 32 33 35 35 35 35 37 37 33 22 22 22 22 30 24 19 16 14 14
12 19 27 28 28 28 24 13 15 16 25 25 30 30 28 33 33 35 35 37 37 35
33 32 32 32 32 27 24 16 16 14 14 14 27 21 30 30 30 23 23 25 28

```

```

28 33 33 33 32 32 30 30 27 15 15 15 19 21 21 12 12 12 22 12 12 14
19 19 19 24 30 32 32 32 32 33 33 33 35 30 30 30 32 32 28 28 28 31
32 23 24 21 22 23 28 23 25 20 22 12 12 11 11 11 16 16 11 11 11 22
18 18 17 20 21 21 20 21 24 16 16 16 16 16 17 12 11 11 11 12 11 11
11 18 20 20 22 25 28 28 21 22 23 27 25 18 18 18 20 28 20 12 12 12
12 12 11 11 11 11 11 11 11 11 12 10 25 12 12 12 16 16 21 29 22 19
21

```

```

>GWO0T3C01A0IV6 length=255 xy=0298_3604 region=1
run=R_2011_01_26_16_16_41_
GAGCGGGAAGGATTTTCGTGCCAACGGCACTCGTACACGAGTTGTTTCGGAACCTGAGCACA
ACGTTATATCGTTTTGTTAGGAAAAGTGAAGGTGTGTTCGGCGGCAGTTGCTTTTCGGGTAGC
TGTTTCGACTGGCCATAACTGAAAAGGGCAACAGAGAAAATAGAAAAAAGAGGAGATGAG
CAGATGGAGGAAGACGGTCAATGTAGCGCAGATTAAGTCTCGAAACTCAAACGAACGAC
AAAAAAAACAGAGAA

```

```

>GWO0T3C01A0IV6 length=255 xy=0298_3604 region=1
run=R_2011_01_26_16_16_41_
40 40 40 40 34 34 34 34 40 40 40 40 30 30 30 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 21 20 20 20 25 30 40 40 37 38 38 37 40 40 40 40 31 31 27
27 31 33 29 17 14 13 13 13 25 18 22 34 18 18 13 14 14 24 32 36 36
36 36 36 40 40 40 40 40 40 40 40 40 40 40 40 40 40 38 38 38 40 39
39 39 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 28 28 28 25 18 18 20 20 40 38 36 36 36 38
36 37 31 31 30 33 25 36 16 18 18 17 17 17 17 40 21 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 38 38 38
40 34 34 34 38 38 40 40 40 40 40 40 40 40 40 40 25 23 23 26 36 38
38 28 29 19 19 19 25 34 40 36 36 36 40 40 34 37 40 40 40 40 15 15
15 15 15 15 15 26 12 25 23 32 28 28

```

```

>GWO0T3C01B2EN0 length=256 xy=0730_1582 region=1
run=R_2011_01_26_16_16_41_
GTGCAAAATCGCTCGTATGATTTGGGTATCGGAGCGAAAGACTCATCGAACCACTAGTAG
CTGGTTCACATCGAAGTTTCCCTCAGGATAGCTGGTGCTAGTAGAAGTATTGTGCGGTAA
AGCAAATGATTAGAGGCATTGGTGTCTTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TTTCGACCTATTCTCTAAACTTTAAATGTGCAACAAACCGTGCCTTAGCCAACCTGCAAGG
ACCGGAAAGAAAAGAA

```

```

>GWO0T3C01B2EN0 length=256 xy=0730_1582 region=1
run=R_2011_01_26_16_16_41_
37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
30 30 30 40 40 39 39 39 40 40 40 40 39 39 39 40 40 40 40 40 40
40 40 40 40 40 40 40 40 35 35 36 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35
35 35 35 35 33 32 25 20 18 18 18 32 33 25 25 25 33 35 35 37 37 37
37 37 37 37 37 37 37 37 37 37 35 20 19 19 24 27 23 25 12 11 11 11
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
11 11 11 11 27 27 27 27 30 29 29 30 31 33 33 33 20 20 16 16 16 26
32 27 26 26 16 14 14 21 27 23 25 25 23 19 19 19 17 17 22 30 30 29
24 23 23 23 23 23 19 19 19 19 21 21 23 24 24 24 23 23 23 23 27 27
27 26 28 17 17 17 16 14 14 14 14 27 12 11

```

```
>GWO0T3C01CL3IZ length=460 xy=0954_2761 region=1
run=R_2011_01_26_16_16_41_
AGAAAACAACCGTGATTCTCTCAGTGAGCGGCGAGCGAAGAGGGAACGAACCTCGTTGCCG
AATCGGGTCTTACAAGGCCTTGAGTTGTGGCACAAATATCATTTTCGTGGTTGGTGCAGGA
GTGGCGAAATTCAAGCGCAAAGCAAACCCGTTGCTGAATACAACCCTTCATGTGAGTATT
GAGCCAAAGAAGGTGTTAGCCCATTGAGCCATAAACCTGAGCGCCTCATGAACTGTATTA
AGCGAGTAGCACTGTTTGGGAATGCAGTGTCAAGTTGGCAGGTATTTGTCTGCCAAAGTT
AAATACAGAGCGGGAAGACGGCGTCGAAAAGCAATCTAGAAACATAACCAACACACGAG
CACGCACGCGCGCGCGCACGCAGCATAGACAAGCAACACGACGAGAACAACAATAAAA
CAAAAAAAAAAAGAAAAGAACAAGACCAAGGAAACGAAAT
```

```
>GWO0T3C01CL3IZ length=460 xy=0954_2761 region=1
run=R_2011_01_26_16_16_41_
27 27 15 16 16 18 20 22 22 28 28 28 35 35 39 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 31 31 31 31 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 36 35 35
35 37 37 37 37 37 37 35 35 35 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 35 29 21 21 21 30 35 35 35
30 21 21 21 35 35 35 33 33 33 33 33 33 25 20 20 20 20 32 32 32
32 32 32 33 33 35 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 25 25 25 38 38 38 35 35 37 37
37 37 37 37 35 35 35 33 33 33 32 32 30 25 25 25 32 32 32 32 27 28
28 25 19 25 32 32 25 25 25 25 27 32 23 25 19 19 23 32 25 32 32 32
32 32 33 30 30 30 35 32 23 30 30 32 32 32 32 32 32 32 24 26 25
25 33 25 25 25 25 24 28 27 27 21 21 21 25 14 13 13 13 13 13 26 26
27 23 21 21 21 23 17 18 18 23 29 25 26 25 27 31 27 23 23 23 29
29 29 33 33 31 29 29 29 27 23 23 23 23 23 23 23 26 25 17 17 18 23
27 27 27 25 31 29 29 29 29 29 29 27 31 29 23 23 17 18 14 14 18 17
23 18 21 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12
21 23 23 25 23 21 21 19 19 25 25 17 15 15 17 15 11 11 11 11
```

```
>GWO0T3C01EN1ST length=272 xy=1797_0139 region=1
run=R_2011_01_26_16_16_41_
ATCATTTTCGTGGTTGGTGCAGGAGTGCCGAAAATTCAAGCGCAAAGCAAACCCGTTGC
TGAAATACAACCTTCATGTGAGTATTGAGCCAAAGAAGGTGTTAGCCCATTGAGCCATAA
AACCTGAGCGCCTCATGAACTGTATTAAGCGAGGAGGATTGAGCTGTAGGGAACGGTGGT
TTCAAAGCAGAGTGCAGGATGAGAAAGGTACAGCCGTTGCTCCGTGCATGTGAGCGTTT
CAGACGTAAAAAAAAACAAAAAGAAAACACCT
```

```
>GWO0T3C01EN1ST length=272 xy=1797_0139 region=1
run=R_2011_01_26_16_16_41_
36 36 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 39 39 37 35
35 33 33 33 32 30 27 16 16 16 16 27 27 25 24 24 27 28 28 28 11 11
11 11 22 25 13 13 13 11 11 11 16 28 24 24 28 28 28 19 13 13 13 19
14 22 17 16 12 12 12 12 17 19 19 22 21 27 25 24 27 28 31 32 30 30
30 29 29 26 23 25 25 21 21 23 23 24 23 28 31 32 32 28 28 28 30 28
28 25 25 24 26 24 19 14 13 13 13 13 13 13 21 22 22 28 28 28 29 29
30 29 32 28 28 28 32 31 30 30 27 27 27 29 23 23 24 30 30 31 29 29
32 24 24 25 32 33 32 28 28 28 32 32 32 30 25 25 25 33 35 35 32 29
30 25 25 25 25 30 35 30 30 28 28 28 30 30 30 30 32 33 33 33 29 29
30 31 25 25 25 28 25 25 25 25 35 26 30 30 30 37 37 35 35 37 37 37
34 34 26 26 25 25 28 30 30 30 30 30 27 23 24 23 16 15 15 13 21 25
28 28 28 17 13 11 11 12 12 12 12 12 12 20 12 12 12 12 12 27 11
11 14 14 37 24 33 28 27
```

```
>GWO0T3C01C02EM length=82 xy=1125_0732 region=1
run=R_2011_01_26_16_16_41_
GCTGGTTCACATCGAAGTTTCCCTCAGGNTAGCGTGGTGCTAGTAGAAGTATTGTGCGGT
AAGCAAAAAAAAAATGATTAGAG
```

```
>GWO0T3C01C02EM length=82 xy=1125_0732 region=1
run=R_2011_01_26_16_16_41_
32 30 25 25 25 30 30 28 32 32 30 30 28 28 27 21 18 13 13 13 13 13
13 12 13 25 22 22 0 22 22 19 19 18 19 19 23 21 21 22 22 22 25 24
27 24 13 13 14 19 14 13 13 25 26 20 19 16 16 22 20 22 21 19 11 11
11 11 11 11 11 11 11 14 22 22 12 12 14 19 14 14
```

```
>GWO0T3C02JPXBS length=459 xy=3868_2918 region=2
run=R_2011_01_26_16_16_41_
AGTCAAACAGACGCGCAAGGGAGGAGATATGGTTGCCAGTACTTCTCTCGTACTGAAAGG
GAGATGCAAAATGCATCTGTGGTGTGCACACGTCAAATCCGTTTGATTTGCCCAACACCG
ACCGGCCCTCCGTGGGTTGCGTTGGAGTGTGCCTGTTTGGACCCGAAAGGTGGTGAACCTA
TGCCTGAACAAGGTGAAGCCCGGGAAACCCAGGTGGAGGCCTGTAGCGATGCTGACGTG
CAAATCGCTCGTATGATTTGGGTATCGGAGCGAAAGACTCATCGAACACCTAGTAGCTG
GTTACATCGAAGTTTCCCTCAGGATAGCTGGTGCTAGTAGAAGTATTGTGCGGTAAAGC
AAATGATTAGAGGCATTGGTGTCTTAGAATATCGACCTATTCTCAAACCTTAAATGTGC
AATGCAAGAACAAAAAAAAACTCGAAAAACAAAAACGAA
```

```
>GWO0T3C02JPXBS length=459 xy=3868_2918 region=2
run=R_2011_01_26_16_16_41_
40 40 40 40 21 21 21 31 40 40 40 40 40 40 40 40 40 40 39 39 39 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 37 37 37 39 39 39 40 40 40 40 40
40 34 34 34 34 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 31 31 31 40 40 40 40 39 39 39 40 40 39 39 39 40
39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40
40 39 37 37 37 37 40 40 40 40 40 39 39 40 40 40 40 40 40 40 40 40
40 39 39 39 37 37 37 37 35 35 37 28 28 28 37 37 38 37 37 37 40 40
40 40 40 39 39 37 37 37 40 37 37 37 40 40 40 40 40 40 40 40 40 40
39 39 39 34 34 34 34 39 39 39 27 22 22 40 40 40 37 37 37 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39 40 39
39 39 40 40 40 40 40 40 40 40 40 40 40 39 20 20 20 39 39 39 40 40
40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40
35 35 21 21 21 21 26 31 33 30 30 27 24 26 26 26 40 31 31 31 31 31
```

```

40 40 40 40 40 40 40 40 37 37 37 40 35 35 35 35 35 35 31 24 26 25
27 27 32 26 25 25 25 19 14 14 14 28 32 33 33 33 31 28 19 19 19 28
28 33 33 33 28 25 25 30 30 33 28 28 28 32 24 24 25 28 28 28 28 33
30 30 32 30 35 36 36 28 24 19 19 19 24 33 27 27 22 14 14 14 19 19
17 17 17 17 20 16 16 15 11 11 11 11 16 11 11 11 11 11 11 11 11 16
20 19 20 13 14 14 14 16 21 10 10 10 10 10 10 19 20 8 8

```

```

>GWO0T3C02JLGXG length=433 xy=3817_3938 region=2
run=R_2011_01_26_16_16_41_
CGGCCGTAACGCCTTTTCAACTCACGGCCTCTAGGAATGAAGGAGGGTAGTTTCGGGGGAG
AACGTACTGGGGCGTCAGAGGTGAAATTCTTAGACCGCACCAAGACGAACTACAGCGAAG
GCATTCTTCAAGGATACCTTCTCAATCAAGAACCAAAGTGTGGAGATCGAAGATGATTA
GAGACCATTTGTAGTCCACACTGCAAACGATGACACCCATGAATTGGGGATCTTATGGGCC
GGCCTGCGGCAGGGTTTACCCTGTGTGTCAGCACCGCGCCCGCTTTTACCAACTTACGTATC
TTTTCTATTTCGGCCTTTACC GGCCACCCACGGGAATATCCTCAGCACGTTTTTCTGTTTTT
TCACGCGAAAGCTTTGAGGTTACAGTCTCAGGGGGGAGTACGTTACACGCAACAAAACA
GCAAGAAAAAAA

```

```

>GWO0T3C02JLGXG length=433 xy=3817_3938 region=2
run=R_2011_01_26_16_16_41_
40 40 40 38 38 38 40 40 40 40 40 40 40 34 34 34 34 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
21 21 21 40 40 40 40 40 40 26 26 26 26 26 40 31 40 40 40 40 40
39 38 21 21 21 21 38 29 38 40 40 40 40 39 39 40 40 39 39 39 40 40
40 31 31 31 38 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 29 29 29 38 40 40 40 40 40 40 40 40 40 40 40 40 40 39 39 40 40
40 40 40 39 40 40 40 40 40 40 40 40 40 40 40 40 40 39 40 40 40
40 40 40 40 40 39 39 39 40 40 39 40 40 40 40 40 38 33 33 33 40 40
40 40 40 40 34 34 33 33 40 40 40 40 40 40 40 39 39 39 40 40 40
40 40 40 40 40 40 40 40 40 39 39 39 28 28 28 40 39 39 39 40 40
40 40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 34 34 34 34 40
38 38 38 38 38 38 37 35 35 39 36 36 35 28 20 20 20 20 35 24 35 35
35 38 38 39 39 39 39 39 39 40 40 40 38 38 38 38 39 33 33 33 35 35
33 33 33 39 39 40 39 39 39 40 38 38 38 38 32 32 32 32 28 28 28 28
35 24 40 17 22 22 22 17 16 39 39 39 39 39 39 36 36 33 30 31 20 20
20 31 31 33 33 36 36 36 32 28 30 30 30 28 21 22 11 11 11 11 11
11 9 18 15 25 23 23 28 30 26 19 20 20 26 25 20 20 20 23 20 19 19
23 12 20 20 19 15 17 11 11 11 11 11 11 11 11 11

```

Texto Suplementar S2. Reads de mRNAs presentes nas bibliotecas transcriptômicas apresentam caudas homo- ou heteropoliméricas nas extremidades 3'-terminal. O conteúdo das seqüências e os respectivos scores de qualidade estão representados.

```
>GKLMG1101B8T40 length=546 xy=0803_2550 region=1
run=R_2010_07_15_13_30_27_
GCCGTGCAGATCTCGATCGCGAACGTTGGCTCGAACGGCGTGATCGACGGCACGACGACG
ACGCTGTGCATTGCTGGCTACCTGCGCTCGCAGGGCGAGTCGGACCACGCGATCAACCAC
ATCGCGATCGACCGCAGAATCGTGCATCAAGACCGGCAAGCCGAAGCGCGCTCCAAG
TCGAAGTCGAAGAAGCCCCGCCGCGAGGGGCGCTGCCGCTGGTGGTGTGCTGCTGCGCAG
AAGGGGGCTCGCCCGTCTGCCCAGAAGGGAGCCCGCCGCTGCACAGAAGGGGGCTCGC
CCGCTGCGCAGAAGGGAGCTGCCCGACCCGCAAGAGCCGTGCTTAGAGGAAGAGCGGG
AAAGCTCAAATGAAAAGAAAAAAGGGTAAAAACGAGAAGACCCACAGCAGTAAAAGAGGA
GGGCTGCTGGTTTCCCTCTTTGACTAGTGGTAGAAAAAAGAGGGGCCGTGATTTTC
TTTACCTTTAACTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA
TAAAAA
```

```
>GKLMG1101B8T40 length=546 xy=0803_2550 region=1
run=R_2010_07_15_13_30_27_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 39
39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
20 18 18 18 18 32 32 25 26 16 16 16 16 25 27 35 35 29 28 30 30 32
32 32 32 32 31 36 27 26 26 26 38 25 16 16 23 27 25 39 39 40 40 40
40 40 40 40 28 28 28 28 28 40 40 39 37 32 32 32 35 40 40 40 40 40
40 35 36 36 40 40 39 32 32 32 30 27 27 32 32 33 31 31 38 40 25 25
21 27 15 21 21 27 32 35 39 40 38 38 35 32 32 32 36 32 27 27 27 36
36 36 36 36 34 29 24 24 16 14 14 22 32 29 29 26 22 22 22 22 14 14
14 14 19 9 9 9 9 9 12 17 17 21 30 20 19 22 22 24 35 36 22 39 3333
33 28 30 30 33 33 25 30 26 26 26 26 26 21 21 21 21 24 14 24 24 23
18 18 18 24 22 22 18 18 17 17 15 15 18 20 20 20 15 15 22 23 23 23
24 18 18 16 17 20 17 17 14 11 11 11 11 11 11 11 11 11 11 11 7 13
14 13 13 15 15 12 22 15 15 16 18 16 15 16 16 22 12 12 14 24 16 16
12 12 12 17 17 20 15 19 12 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 12 13 7 7 7 7 7 7 7 7 7 7 7 7 7 7 14 20 7
7 7 7 7 14 8 8 8 8 8 14 8 8 8 8 8
```

```
>GKLMG1101DBLE9 length=265 xy=1245_0419 region=1
run=R_2010_07_15_13_30_27_
ACAACGCGTCCAAGACGTACCCCTGGCTGCTGGTGCCTTGAAGAAGGGTGGATACGTGT
GCATCAACGGCCGTCCGTGTAAGGTGATTGATCTGTCCGTGTCCAAGACCGGCAAGCATG
GTCATGCTAAGGTGAGCATTGTTGCGACCGACATCTTCACTGGCAACCGCCTCTCTAGAG
GACCCCGCAGGAGGACGGTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA
ANAAAAAANAAAAAANAAAAA
```

```
>GKLMG1101DBLE9 length=265 xy=1245_0419 region=1
run=R_2010_07_15_13_30_27_
36 36 31 31 31 37 37 37 37 37 37 39 37 37 39 39 39 37 32 22 22 20
20 20 28 28 28 32 30 31 31 31 33 33 33 32 32 32 35 35 31 30 30 35
32 32 30 34 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
```

```

37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 36 36
36 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 32 28 28 28 31 31
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 21 21 21 20 32 26 32 32 31 31 29 23 23 25 24 24
33 12 12 12 12 12 12 12 12 12 12 12 10 10 10 10 10 10 10 10 10
10 10 10 10 10 0 10 10 10 10 10 10 10 10 12 12 12 12 12 12 0
10 10 10 10 10 10 0 12 15 15 15 12 11 11 0 11 12 12 12 12 10
10

```

```

>GKLMG1102G60U9 length=383 xy=2833_1171 region=2
run=R_2010_07_15_13_30_27_
ACTATATCCACAGGGAGAAGCGGATGAAGAACGCTAATGTGGTGACGCTGTTCCGACTGC
GTAGCTGCCCGGGCGCCCTTTTCTACAGATGCTACGTCAAGTCGCGCATATGCGCACCT
CCCCTCCCTCTCTTCTCCCTCGCGCCGGTATTGAAGCTCTCTATCGACGGCTTGTTTCG
TGTGCACGTGGCCATGTGTGCCTGTCTTTCCAGGCATTTTATTGTTGCTTCTCTGCGTC
TTTTCTCCGCCCCCACCTCTTGAACGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAA

```

```

>GKLMG1102G60U9 length=383 xy=2833_1171 region=2
run=R_2010_07_15_13_30_27_
37 37 37 31 30 30 37 37 37 37 37 37 30 30 30 36 37 37 37 39 40 40
39 39 39 40 40 40 40 40 40 40 40 40 39 39 40 39 40 40 40 39 39 40
40 40 40 40 40 40 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 35 35 28 28 28 28 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 35 35 37 37 37 37 37 37 32 32 32 32 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
16 19 14 16 30 29 29 30 27 23 25 25 30 37 37 37 25 25 25 33 24 25
22 19 23 14 17 17 22 11 11 11 11 11 11 18 10 10 10 12 13 13 16 16
28 11 12 12 14 16 19 20 12 12 12 12 12 12 12 12 12 12 12 12 12
12 12 12 12 0 10 10 10 10 10 10 10 10 10 10 10 8 8 8 8 8
8 8 8 8 8 0 8 10 10 10 10 10 10 10 10 10 8 8 0 10 10 10 10
10 10 10 10 8 8 0 8 8 10 10 10 10 10 10 10 10 8 8 0 8 8
10 10 10 10 10 10 7 7 0 7 7 10 10 10 10 10 10 7 7 0 7
7 7 7 7 7 7 7

```

```

>GKLMG1102INSWN length=303 xy=3434_1925 region=2
run=R_2010_07_15_13_30_27_
GCCGCCAGATGGTGCTGCAGGACAAGACGAAGTTTGGCTCACCCAAGTACCGCCTTGTTG
TGCGCACGACGAACAAGGACATCATTCGCGAGATCGTGCAGGCGAAGATCGCCGGCGACG
AGGTGCTGATGGCTGCGTACGCGCACGAGCTGCCTGCGTTCGGGATTGAGCACGGCCTGA
CGAACTACGCTGCGGCGTACGCGACGGCCTGCTGCTGGCGCGCCGACGCTGGCGAAGT
TGGGCATCGCGGACAAGTTCCAGGGCGAACGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAA

```

```

>GKLMG1102INSWN length=303 xy=3434_1925 region=2
run=R_2010_07_15_13_30_27_
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 39 39 39
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40

```

```

40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 39 39 39 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 38 21 21 21 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 38 38 38 40 40 40 39
39 39 40 40 40 40 40 40 38 38 38 38 38 40 40 40 40 40 38 40 31 25
18 18 13 13 13 12 17 11 11 11 11 11 11 11 11 11 11 11 11 11 0 10
10 10 10 10 10 10 10 10 0 10 10 13 13 13 13 10 10

```

```

>GKLMG1102F1B2V length=369 xy=2358_1749 region=2
run=R_2010_07_15_13_30_27_
GACCACGCAAGCGTCGGCGTTGCTCACTCGTCGTCGTCGGCCTCTTCTTTAGAGACCGAG
ACGCCACAGCAGGCAAACAGGACAAGCGATTACACTAGCAATCCCGGTGAGCGAGAGAAG
GAATGCTGCCGAAGCGGTCTGTGTATCCAATTTGTGCGCCGACGTCTTCTGCGCGGCC
CTGGCGGTCAAACGCGCTGGCACCGCTCCGTACTGTGAGCCGAGTGTGCGCCAGAGCCCA
TCCGCCTCCTTTGCCAGCATGCCACGCCATATGGAAGTAAAAAAAAAAAAAAAAANAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
ANAAAAA

```

```

>GKLMG1102F1B2V length=369 xy=2358_1749 region=2
run=R_2010_07_15_13_30_27_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 39 39 39 40 40 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 35 23 23 23 27 37 37 37 37 37 34 35 35 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 35 35 35 35 35
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
35 37 35 38 38 38 38 35 35 33 33 33 32 30 12 12 12 12 12 12 12 12
12 12 12 0 10 10 10 10 10 10 10 10 0 10 10 10 10 8 8 8 8 0
10 12 12 12 12 12 10 10 0 12 12 12 12 12 12 10 8 0 8 8 8 8
8 8 8 8 0 8 8 8 8 8 8 8 0 8 8 8 8 8 8 8 8
0 8 8 12 10 10 10 7 7 0 7 7 7 7 7 7 7

```

```

>GPUPS7E02F1JNG length=342 xy=2360_3370 region=2
run=R_2010_10_21_16_08_57_
ACGACACGTATCCACGATGGACTACGCCACCAACGCGATGGCCATCACCAACAAGGCGAA
GGTGAACAAAAGTTTCAAGAGACTTTGAGTATAGTGATCTCCTAAAGATGCAAAAGTACC
TTGAGGTGTCTGCAGCGCGTGAAACGAGTCAGATTGAAAGCTACGTCGCTCAAACGGCTG
AGTTGAAGGCGGAGATCGAGGCACTTAATCAGGAGCTGCTGACGCTGAACGATTACAGC
TCGAGGCTGGGTTGCAGGAGGCACTTTCCGAGCGGGATGCGCTTGTATCGGAAATGGTGT
CGAAGCTTGCTGGCGTTGGTCTAAAGGGCGGAAAAAAAAAAAAA

```

```

>GPUPS7E02F1JNG length=342 xy=2360_3370 region=2
run=R_2010_10_21_16_08_57_
37 37 37 37 37 37 37 37 37 37 37 37 40 40 40 40 40 40 40 40 39 39
39 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
40 40 40 40 40 40 40 40 37 37 36 36 36 37 37 37 37 37 37 35 30 30
30 26 26 26 25 21 21 21 21 32 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 33 33 33 33 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37
37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37 37

```


Soluções e Tampões

Água DEPC

- 1) 1 mL de DEPC em 1000 mL de H₂O ultra-pura.
- 2) Deixar overnight sob agitação, armazenar em Erlenmeyer e autoclavar por 40 minutos para remover os traços de DEPC.

Solução 0.9% NaCl

- 1) 9 g de NaCl em 1000 mL de H₂O destilada.
- 2) Autoclavar por 15 minutos.

Cicloheximida

Preparar solução estoque de 10 mg/mL com H₂O DEPC autoclavada.

Heparina (20 mg/mL)

Preparar solução estoque de 20 mg/mL com H₂O DEPC autoclavada.

Solução de lise NaOH/SDS

- 1) 2.5 mL de NaOH 4 M
- 2) 5 mL de SDS 10% ([]_f = 1%)
- 3) H₂O ultra-pura nuclease-free para completar o volume para 50 mL

Tampão de lise hipotônico

- 1) Preparar e diluir todas as soluções em H₂O DEPC autoclavada.
- 2) Na hora do uso complementar o tampão de lise com cicloheximida, heparina e β-mercaptoetanol. Acrescentar aos 100 mL do tampão de lise: 100 μL de cicloheximida (solução estoque 10 mg/mL), 100 μL de heparina (solução estoque 20 mg/mL) e 34 μL de β-mercaptoetanol. []_f ciclo = 10 μg/mL, []_f heparina = 20 μg/mL e []_f β-mercaptoetanol = 5mM.

	Para 100 mL (1X)	[] _f da Solução de uso
Tris-HCl (PM 121.1) 1 M pH 7.5	1 mL	10 mM
MgCl ₂ · 6H ₂ O (PM 203.3) 1 M	1 mL	10 mM
NaCl (PM 58.4) 5 M	0.2 mL	10 mM
β-mercaptoetanol (14.7 M)	34 μL	5 mM

Nonidet P-40

Preparar solução estoque 10% v/v em tampão de lise hipotônico.

Sacarose 2M

- 1) 68.4 g de sacarose (PM 342.3) em 100 mL de tampão de lise hipotônico.
- 2) Na hora da ultra-centrifugação complementar a sacarose 2M com cicloheximida, heparina e β-mercaptoetanol. Acrescentar aos 2 mL de sacarose 2M: 2 μL de cicloheximida (solução estoque 10 mg/mL), 2 μL de heparina (solução estoque 20 mg/mL) e 0.68 μL de β-mercaptoetanol. []_f ciclo = 10 μg/mL, []_f heparina = 20 μg/mL e []_f β-mercaptoetanol = 5mM.

Tampão NKM

- 1) Preparar e diluir todas as soluções em H₂O DEPC autoclavada.
- 2) Na hora do uso complementar o tampão NKM com cicloheximida, heparina e β-mercaptoetanol. Acrescentar aos 100 mL do tampão de lise: 100 μL de cicloheximida (solução estoque 10 mg/mL), 100 μL de heparina (solução estoque 20 mg/mL) e 34 μL de β-mercaptoetanol. []_f ciclo = 10 μg/mL, []_f heparina = 20 μg/mL e []_f β-mercaptoetanol = 5mM.

	Para 100 mL (1X)	[] _f da Solução de uso
NaCl 5 M	2.8 mL	140 mM
KCl 1 M	0.5 mL	5 mM
MgCl ₂ 1 M	0.15 mL	1.5 mM
Hepes 1 M pH 7.4	1 mL	10 mM

Tampão TAE 10X

- 1) Tris-base: 48.4 g
- 2) Ácido acético glacial: 11.42 mL
- 3) EDTA: 20 mL (0.5 M, pH 8.0)
- 4) Completar o volume para 1L

Para preparar TAE 1X diluir a solução estoque 10 vezes e acertar o pH para 8.3

Meio Grace's base

- 1) Grace's: 45.7g
- 2) Dissolver o pó de Grace's em cerca de 900mL de H₂O ultra-pura
- 3) Acertar o pH para 6.5
- 4) Completar o volume para 1L
- 5) Filtrar em membrana Millipore 0.22 μm estéril e armazenar a 8°C

OBS.: não acrescentar NaHCO₃

Meio Grace's completo

- 1) Soro fetal bovino: 10 mL
- 2) Penicilina: 1mL
- 3) L-glutamina: 1mL
- 4) Meio Grace's base: quantidade suficiente para (q.s.p.) 100mL
- 5) Acertar o pH para 6.5
- 6) Filtrar em membrana Millipore 0.22 μm estéril e armazenar a 8°C

OBS.: não acrescentar NaHCO_3

Meio RPMI

- 5) RPMI: 10.4g
- 6) NaHCO_3 : 2.0g
- 7) Dissolver o pó de RPMI em cerca de 900mL de H_2O ultra-pura
- 8) Acertar o pH para 4.0 (determinados componentes do RPMI só se dissolvem em pH ácido)
- 9) Acertar o pH para 7.2
- 10) Acrescentar o NaHCO_3
- 11) Acertar o pH para 7.2
- 12) Completar o volume para 1L
- 13) Filtrar em membrana Millipore 0.22 μm estéril e armazenar a 8°C

Meio DMEM

- 1) DMEM: 13.4g
- 2) NaHCO_3 : 3.7g
- 3) Dissolver o pó de RPMI em cerca de 900mL de H_2O ultra-pura
- 4) Acertar o pH para 7.2
- 5) Acrescentar o NaHCO_3
- 6) Acertar o pH para 7.2
- 7) Completar o volume para 1L
- 8) Filtrar em membrana Millipore 0.22 μm estéril e armazenar a 8°C

Gel de agarose 1.2%

- 1) Dissolver 0.6 g de agarose em 50 mL de tampão TBE pH 8.3
- 2) Aquecer a mistura até ferver para solubilizar a agarose.
- 3) Após o resfriamento colocar na cuba e aguardar a polimerização.

Ribonucleoside Vanadyl Complex (inibidor de RNases)

- 1) A solução estoque está na concentração de 200 mM (New England BioLabs, Ipswich/MA, EUA).
- 2) Diluir o inibidor para concentração final de 10 mM em tampão de lise celular.

L-Glutamina (solução estoque 200 mM)

- 1) L-glutamina: 2.924g
- 2) 100mL de H₂O ultra-pura
- 3) Esterilizar com filtro de seringa 0.22 µm
- 4) Distribuir alíquotas em eppendorfs de 1 mL e congelar a -20°C

Meio LB-ágar

- 10 g de bacto-triptona

- 5 g extrato de levedura

- 5 g de NaCl

- 1) Dissolver todos os reagentes em 800 mL, ajustar o pH para 7.5 com NaOH e adicionar H₂O destilada para 1000 mL
- 2) Adicionar 4.5 g de ágar e autoclavar por 20 minutos
- 3) Permitir que o meio esfrie até 55°C
- 4) Adicionar 2 mL de solução estoque de ampicilina (50 mg/mL) para uma concentração final de 100 µg/mL
- 5) Misturar levemente e espalhar nas placas

PBS

1) 10X

- NaCl: 81.8 g

- Na₂HPO₄.7H₂O: 19.8 g- NaH₂PO₄.H₂O (anidro): 3.13 g- q.s.p. 1 L de H₂O ultra-pura

2) 1X

- diluir 100 mL de PBS 10X em q.s.p. 1 L de H₂O ultra-pura

- acertar o pH para 7.2

Tampão TBE 5X

Reagente e [] _f da Solução	Quantidade para 1 L do tampão
Tris Base – 445 mM	54 g
Borato – 445 mM	27.5 g de ácido bórico
EDTA – 10 mM	20 mL de solução 0.5 M (pH 8.0)
H ₂ O nuclease-free	q.s.p 1 mL

Para preparar TBE 0,5X diluir a solução estoque 10 vezes e acertar o pH para 8.3

Tampão GTE

- 1) 23 mL de glicose 20%
- 2) 10 mL de EDTA 0.5 M pH 8.0
- 3) 13 mL de Tris-HCl 1 M pH 7.4
- 4) H₂O para completar o volume para 500 mL
- 5) Autoclavar e guardar em alíquotas no freezer -20°C

Penicilina (solução estoque 6 mg/mL)

- 1) Penicilina G: 314 mg
- 2) 50 mL de PBS 1X
- 3) Esterilizar com filtro de seringa 0.22 µm
- 4) Distribuir alíquotas em eppendorfs de 1mL e congelar a -20°C

Ampicilina (solução estoque 5 mg/mL)

- 1) Ampicilina: 500 mL
- 2) 100 mL de H₂O ultra-pura
- 3) Esterilizar com filtro de seringa 0.22 µm
- 4) Distribuir alíquotas em eppendorfs de 1mL e congelar a -20°C