Received: 3 April 2017 | Accepted: 14 September 2017

DOI: 10.1111/jbg.12299

ORIGINAL ARTICLE

WILEY Journal of Animal Breeding and Genetics

Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle

R.R. Mota[1,2] | S.E.F. Guimarães[2] | M.R.S. Fortes[3] | B. Hayes[4] | F.F. Silva[2] | L.L. Verardo[2] | M.J. Kelly[4] | C.F. de Campos[2] | J.D. Guimarães[5] | R.R. Wenceslau[6] | J.M. Penitente-Filho[2] | J.F. Garcia[7] | S. Moore[4]

[1]TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech Faculty, University of Liège, Gembloux, Belgium

[2]Department of Animal Science, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[3]School of Chemistry and Molecular Biosciences, the University of Queensland, Brisbane, Qld, Australia

[4]Queensland Alliance for Agriculture and Food Innovation, the University of Queensland, Brisbane, Qld, Australia

[5]Department of Veterinary Medicine, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil

[6]Animal Science Institute, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

[7]Department of Support, Health and Animal Production, Faculdade de Medicina Veterinária de Araçatuba, UNESP – Universidade Estadual Paulista, Araçatuba, São Paulo, Brazil

Correspondence
R.R. Mota, TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech Faculty, University of Liège, Gembloux, Belgium.
Email: rrmota@ulg.ac.be

Funding information
CNPq; FAPEMIG; CAPES; INCT—Ciência Animal

Summary

We performed a genome-wide mapping for the age at first calving (AFC) with the goal of annotating candidate genes that regulate fertility in Nellore cattle. Phenotypic data from 762 cows and 777k SNP genotypes from 2,992 bulls and cows were used. Single nucleotide polymorphism (SNP) effects based on the single-step GBLUP methodology were blocked into adjacent windows of 1 Megabase (Mb) to explain the genetic variance. SNP windows explaining more than 0.40% of the AFC genetic variance were identified on chromosomes 2, 8, 9, 14, 16 and 17. From these windows, we identified 123 coding protein genes that were used to build gene networks. From the association study and derived gene networks, putative candidate genes (e.g., *PAPPA, PREP*, *FER1L6*, *TPR*, *NMNAT1*, *ACAD10, PCMTD1, CRH*, *OPKR1*, *NPBWR1* and *NCOA2*) and transcription factors (TF) (*STAT1, STAT3, RELA, E2F1* and *EGR1*) were strongly associated with female fertility (e.g., negative regulation of luteinizing hormone secretion, folliculogenesis and establishment of uterine receptivity). Evidence suggests that AFC inheritance is complex and controlled by multiple loci across the genome. As several windows explaining higher proportion of the genetic variance were identified on chromosome 14, further studies investigating the interaction across haplotypes to better understand the molecular architecture behind AFC in Nellore cattle should be undertaken.

KEYWORDS
beef cattle, gene function, single-step

1 | INTRODUCTION

Although Nellore cattle are well adapted to Brazilian climatic conditions and production systems, low reproductive efficiency is still the major reason for high culling rates of beef cows (Utsunomiya et al., 2014). The reproductive trait that is most often used to evaluate female fertility in beef cattle breeding programmes is the age at first calving (AFC). AFC is easily measured, expressed in almost all cows under genetic evaluation and directly affects the herd productivity. Currently, the availability of dense single nucleotide polymorphism (SNPs) markers has allowed for

genome-wide association studies (GWAS) of economically important traits in domestic animals. In terms of fertility traits, some results have been published for beef cattle; for example, Fortes, Li, Collis, Zhang, and Hawken (2013) reported *IGF1* pathway genes related to cattle puberty in Brahman cattle. However, information on GWAS for AFC while combining pedigree and genomic information (Wang, Misztal, Aguilar, Legarra, & Muir, 2012) in Nellore cattle is still scarce in the literature.

The characterization of chromosome regions affecting AFC in Nellore cattle may assist in breeding efforts through the identification of candidate genes by promoting a better understanding of their reproductive biology. Candidate gene detection is complex for low-heritable complex traits such as AFC. The challenge is greater when using data from genotyped and non-genotyped animals. We exploited the biological processes based on transcription factors (TF) in gene-TF network analyses to identify the most relevant candidate genes.

2 | MATERIAL AND METHODS

2.1 | Ethical approval

This study was developed using pre-existing datasets. All animal procedures were approved by the Animal Care and Use Committee of the Department of Animal Science from Universidade Federal de Viçosa, Brazil (103/2014-CEUAP; see http://www.pveuq.ufv.br/courses.php for further details).

2.2 | Genotype, phenotype and pedigree data

The DNA of each animal was obtained from blood or semen samples. The genotype data set was comprised of three subsets of Nellore animals. The first one was composed of 2,272 bulls and cows that were genotyped with the Illumina High-Density Bovine SNP chip (777K panel; n = 786,799 SNPs). The second was comprised of 529 cows with the Illumina Bovine 70K panel. The third was composed of 191 cows with the Illumina Bovine 30K panel. Genotype imputation was performed through the *FImpute* software (Sargolzaei, Chesnais, & Schenkel, 2011) as an efficient strategy to derive a unique reference set, that is, 2,992 animals with 777K panel. Genotype quality control was implemented after imputation to remove SNPs that were mapped to autosomes (29 chromosomes) with call rates <0.97, minor allele frequencies (MAF) <0.05 and with a significant deviation from the Hardy–Weinberg equilibrium ($p < 10^{-7}$).

Phenotypic data included AFC records of Nellore cattle that were raised in the North of Minas Gerais state, Brazil (17°51′ South and 40°43′ West longitude). Their hardiness and ability to adapt to the hot climate and historically low rainfall are the main reasons to raise zebu breeds such as Nellore in this part of Brazil. AFC records from 762 cows that were born between 2000 and 2012 were used; they were pasture-raised mainly with *Urochloa decumbens* (*Syn. Brachiaria*) *decumbens*, mineral salt and water ad libitum. The breeding season period is usually between November and February. Seventy per cent of the cows were in between 28 and 32 months old when the breeding season started. AFC overall mean during the evaluation period was 1,199 ± 184.61 days (range: 628–1,821) or 39.9 ± 6.15 months. Contemporary groups (CG) were defined as cows that were born in the same year and season (March to May, June to August, September to November and December to February).

Our studied population is a pseudo-experimental population that was raised under harsh conditions in Brazil under constant reproduction and genetic evaluation. A small number of phenotypes would not allow us to make inferences on a country-wide scale. To overcome this limitation, we added 2,272 genotypes of representative animals from the entire Nellore population in Brazil. These animals were able to link our studied population to the wider Nellore population as long as most of the genotyped animals were Nellore proven bulls.

Pedigree information was recovered from historical breeding records from the Brazilian Association of Zebu Breeders (ABCZ), which was comprised of 6,341 individuals. A total of 4,133 animals remained after pruning the data.

2.3 | Statistical model and linkage disequilibrium analyses for GWAS

The following single-step genomic-BLUP (ssGBLUP) model was fitted:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{W}\mathbf{a} + \mathbf{e} \quad (1)$$

where $\mathbf{Y}$ is the vector of AFC records; $\boldsymbol{\beta}$ is the vector of systematic effects (CG; n = 35 levels), $\mathbf{a}$ is the vector of random additive genetic effects, $\mathbf{X}$ and $\mathbf{W}$ are incidence matrices of systematic and random additive genetic effects, respectively, and $\mathbf{e}$ is the random residual vector.

Genomic and pedigree information were combined via ssGBLUP procedure (Aguilar et al., 2010). The model (1) was implemented using the $\mathbf{H}$ matrix that includes simultaneously SNP markers and pedigree information. Although $\mathbf{H}$ is complex, its inverse can be obtained according to Aguilar et al. (2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{\mathbf{22}}^{-1} \end{bmatrix}, \quad (2)$$

where $\mathbf{G}$ is the genomic relationship matrix using current allele frequencies (VanRaden, 2008), and $\mathbf{A_{22}}$ is a numerator relationship matrix for genotyped animals. The

following distributional assumptions were assumed as follows: $\mathbf{a} \sim N(\mathbf{0}, \mathbf{H}\sigma_a^2)$ and $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$; $\mathbf{I}$ is the identity matrix, σ_a^2 and σ_e^2 are additive genetic and residual variance components, respectively.

Genomic estimated breeding values (GEBV) were obtained via REML using the software *airemlf90* (http://nce.ads.uga.edu/). The GEBVs were further used for SNP effect derivation (Wang et al., 2012). Decomposing the additive genetic effect into genotyped ($\mathbf{a_g}$) and non-genotyped ($\mathbf{a_u}$) animals, SNP effects ($\mathbf{u}$) can be estimated as $\hat{\mathbf{u}} = Z'[ZZ']^{-1}\hat{\mathbf{a}}_\mathbf{g}$, which is the best predictor of SNP effects given animal effects (Wang et al., 2012).

We investigated chromosomal regions where SNP effects were blocked into adjacent windows of 1 Megabase (Mb). Genetic variances were explained per window and estimated through *PostGSf90* (http://nce.ads.uga.edu/) software by calculating the variance explained by the *n* Mb window of adjacent SNPs (segments) with their respective effects as input. This was used to identify potential candidate genes that may affect AFC. SNP markers on the same chromosome were analysed for linkage disequilibrium (LD). The LD between any two loci within the same chromosome (within and across different neighbouring windows) was assessed via the r^2 measure. It is considered to be the most robust LD measure, as it is less dependent on MAF and sample size. The r^2 value was used as an LD measure and calculated using the default of the *Haploview* software (Barrett, Fry, Maller, & Daly, 2005) for every SNP pair.

2.4 | Assessment of gene functional annotation and network analyses

Following Utsunomiya et al. (2014), we selected the largest window variances (>0.40%) as outliers for the most relevant windows that were affecting AFC. To identify putative genes associated with the list of SNP markers located inside the most relevant windows, we used the package Map2NCBI (Hanna & Riley, 2014) of the R software based on the UMD Bos taurus 3.1 assembly of the bovine genome sequence, which allowed us to generate a list of genomic features from the Bos taurus (BUILD.6.1) genome. To provide information regarding the identity and function of genes at adjacent windows, the chromosomal positions from the Ensembl Genome Browser (http://ensemblgenomes.org/) were used. In addition, a Bioconductor package that accesses and retrieves Ensembl data (Entrez IDs, Ensembl gene ID, HGNC symbols and more), R/BiomaRt, was used to download all genes (background genes) from the Bos taurus genome (ORG.MESH.BTA.DB) as well as map features within ±200 kb from the location of the SNP markers.

The biological function of these genes and possible relation to AFC were first investigated using a biological process gene network. For this, the ClueGO plug-in for Cytoscape (Bindea et al., 2009) was used, based on a 1-sided hypergeometric test and Bonferroni correction, to construct a gene network highlighting the biological roles and relations across candidate genes. While aiming to identify the TF related to potential candidate genes, the TFM-Explorer web tool (http://bioinfo.lifl.fr/TFM/TFME/) was used. This web tool takes a set of gene sequences and searches for locally over-represented transcription factor binding sites (TFBS) using weight matrices from the JASPAR vertebrate database (Sandelin, Alkema, Engström, Wasserman, & Lenhard, 2004); it also detects all potential TFBS and extracts significant clusters (region of the input sequences associated with a factor) through score function calculation. The score threshold is given by a *p-value* equal to or smaller than 10^{-3} for each position and for each sequence (Touzet & Varré, 2007). The program default for the analysed promoter region is 2,000–200 bp upstream and downstream, respectively. As the gene transcription start site (TSS) annotations are uncertain in the current assembly for some regions, we compensated both in the 5′ and 3′ directions by applying no restrict ample definitions. Therefore, from this set of genes, excluding the ncRNA genes, we collected sequences of the TSS gene that were 3,000 bp upstream and 300 bp downstream (FASTA format), based on the *Bos_taurus*_UMD_*v3.1.1* assembly (Zimin et al., 2009). This data were used as an input for the TFM-explorer.

The given TF list was fed into Cytoscape (Shannon et al., 2003) using a Biological Networks Gene Ontology tool (BiNGO) plug-in to determine which gene ontology (GO) terms were significantly overrepresented by assuming a default statistical test (binomial test) and multiple testing corrections (e.g., Bonferroni and false discovery rates) with a significance level of 0.05%. The total number of statistical tests in a single analysis might be as much as several hundred. Based on biological processes (e.g., response to growth hormone stimulus and response to fatty acids) and literature reviews, we selected the main TFs related to AFC (key TF) and constructed a gene-TF network. A schematic representation of the work- and dataflow can be seen in Figure 1.

3 | RESULTS

Genetic and residual variance component estimates for AFC were 6,685.00 ± 2,910.70 days and 21,162.00 ± 2,577.77 days, respectively. The heritability estimate (h^2) was 0.24 ± 0.04, which confirmed genetic variability for this trait and for possible selection in this population.

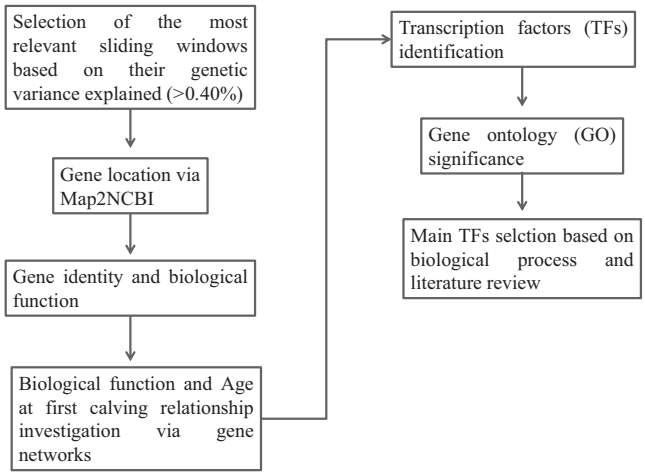

FIGURE 1 A schematic work- and dataflow representation

Single nucleotide polymorphism windows with a 1 Mb length (Figure 2) were built across the genome with an average density of 172 ± 49 SNPs per window. A total of 18 SNP windows that each explained more than 0.40% of the genetic variance were identified. These windows were used to locate candidate genes affecting AFC on BTA 2, 8, 9, 14, 16 and 17. The SNP window (n = 232 SNPs) with the highest proportion of genetic variance explained approximately 3% and was located on BTA14. Nine windows with a threshold that was higher than 0.40% were also located on BTA14, whereas 1, 1, 3, 2 and 1 window(s) were respectively located on BTA 2, 8, 9, 16 and 17. A detailed breakdown of the 1 Mb SNP windows that were used to locate candidate genes is presented in Table 1.

We checked the LD between SNP markers within and across different neighbouring SNP windows (n = 18), which each explained more than 0.40% of the genetic variance. Illustrations of the haplotype blocks in which segments of correlated SNPs are separated by gaps between the gray-scale triangular matrices are presented in Figure S1. Patterns of block structures differed between peaks in the same chromosome and presented with a high LD within each peak, especially on BTA8, BTA9, BTA14 and BTA16.

A total of 3,570 SNP markers were within the relevant adjacent windows and were detected in Bos taurus chromosomes BTA 2, 8, 9, 14, 16 and 17. Potential candidate genes were identified using NCBI and Ensembl sources for gene mapping, and 152 genes (123 protein coding, four snoRNA, 13 snRNA, one miRNA, four rRNA, two misc_RNA and five pseudo genes) were mapped against the major loci that were found for AFC. Some of these SNPs had no identified gene, and 111 genes were mapped within or close to 2,275 SNP markers. Supplementary information linking SNP markers and potential candidate genes is presented in Table S1.

The biological process gene network highlighted a well-related process (negative regulation of luteinizing hormone secretion) using protein-coding genes (Figure 3). From this network, four genes were directly linked: neuropeptides B/W receptor gene (*NPBWR1*, *ENSBTAG00000016159*), nuclear receptor coactivator 2

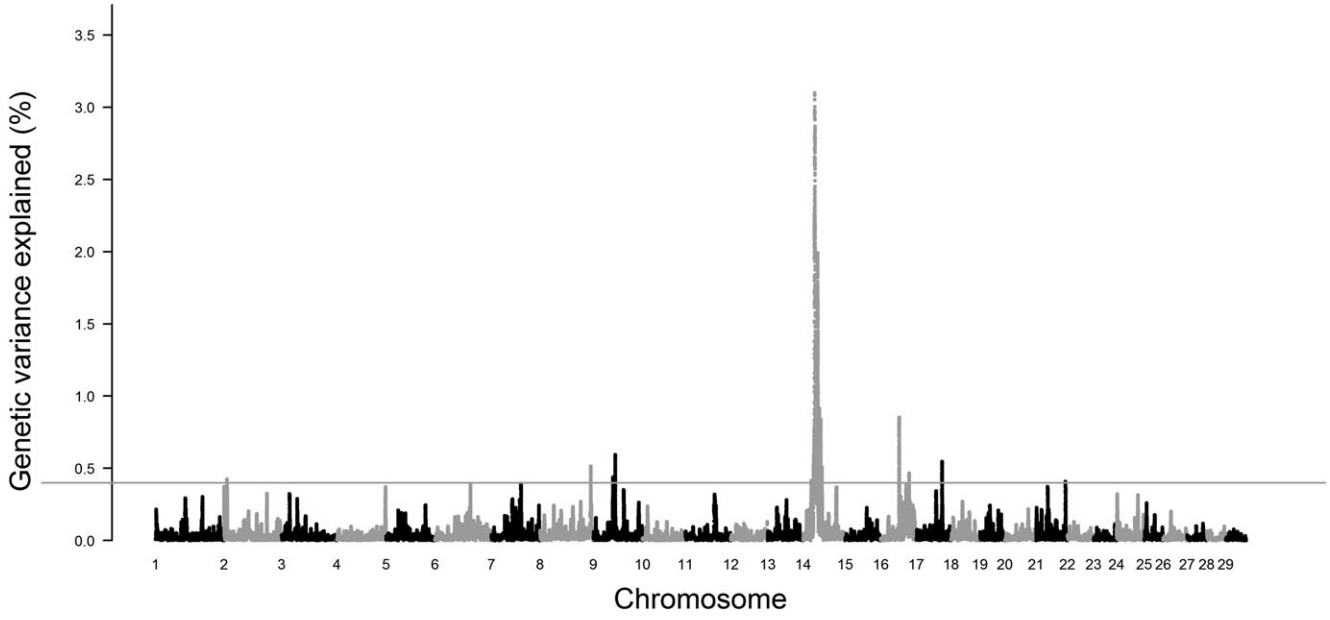

FIGURE 2 Manhattan plots of age at first calving genetic variance explained by SNP windows in Nellore cattle. Each dot represents a 1 Mb SNP window segment. Horizontal gray line represents the adopted threshold (0.40%)

TABLE 1 Relevant single nucleotide polymorphism (SNP) windows detecting major loci explaining variance and potential candidate genes for age at first calving in Nellore cattle

Window start	Window end	Chr	PS (Mb)	PE (Mb)	N. SNP	GVE (%)
BovineHD0200001802	BovineHD0200002125	2	6.17	7.17	178	0.43
BovineHD0800031709	BovineHD0800032090	8	106.27	107.27	219	0.52
BovineHD0900011384	BovineHD0900011657	9	40.97	41.97	178	0.44
ARS-BFGL-NGS-66207	BovineHD0900012314	9	43.31	44.31	235	0.44
BovineHD0900012655	BovineHD0900012880	9	45.62	46.61	131	0.60
BovineHD1400004685	BovineHD1400004991	14	16.54	17.53	214	0.41
BovineHD1400005826	BovineHD1400006161	14	20.39	21.39	217	0.53
BovineHD1400006241	BovineHD1400006529	14	21.66	22.66	196	0.79
BovineHD1400006530	BovineHD1400006869	14	22.66	23.66	232	3.10
BovineHD1400006871	BovineHD1400007165	14	23.67	24.67	198	0.61
BovineHD1400007551	BovineHD1400007803	14	26.20	27.20	183	0.52
BovineHD4100011412	BovineHD1400008347	14	27.84	28.84	237	1.99
BovineHD1400009017	BovineHD1400009324	14	31.25	32.25	216	0.92
BovineHD1400009735	BovineHD1400010082	14	33.91	34.91	231	0.84
BovineHD1400010352	BovineHD1400010608	14	35.96	36.95	186	0.51
BovineHD1600012226	BovineHD1600012468	16	43.94	44.93	147	0.85
BovineHD1600019305	BovineHD1600019652	16	68.23	69.23	194	0.47
BovineHD1700016259	BovineHD1700016502	17	57.29	58.28	178	0.55
Total		N/A	N/A	N/A	3,392	14.52

Chr, chromosome; PS, position start; PE, position end; N. SNP, number of SNPs; GVE, genetic variance explained; N/A, non-applicable.

(*NCOA2*, *ENSBTAG00000020312*), opioid receptor Kappa 1 (*OPRK1*, *ENSBTAG00000000914*) and corticotropin-releasing hormone (*CRH*, *ENSBTAG00000033128*) on BTA14.

From the promoter sequences of the 123 protein-coding genes that were mapped, a total of 18 TFs were identified and analysed to determine which GO terms were significantly overrepresented. The main TFs associated with AFC based on biological processes and literature review (Table 1) were chosen to generate a gene-TF network enabling the identification of putative candidate genes for AFC (Figure 4). Based on the most representative TF (*STAT1*, *STAT3*, *RELA*, *E2F1* and *EGR1*; Table S2), we identified the most likely candidate genes for AFC (e.g., *ORMDL1* on BTA2; *PAPPA* on BTA8; *FIG 4*, *PPIL6*, *PREP* on BTA9; *PCMTD1*, *SDCBP*, *UBXN2B*, *MCM4*, *ARMC1*, *MRPL15*, NSMCE2 and *FER1L6* on BTA14; *APITD1*, *NMNAT1*and *TPR* on BTA16 and *ACAD10* and *PRKAB1* on BTA17).

4 | DISCUSSION

After quality control edits, 399,308 (51%) SNP markers and 4,133 animals remained in the HD panel to impute 30k and 70k panels and to further estimate the genomic relationship coefficients between animals. The number of overlapped SNPs was 13,708 and 53,375 on 30k and 70k panels, respectively, whereas 2,900 (30k) and 3,673 (70k) were excluded.

The estimated population mean for AFC in this study (1,199 ± 184.61 days) agrees with those reported in the literature, which varied from 1,050 to 1,260 days in Brazilian Nellore cattle (Grossi et al., 2008). High AFC values may be due to the harsh climate conditions, as the animals were pasture-raised in a region with historically low yearly rainfall (Figure S2; http://clima1.cptec.inpe.br/evolucao/pt) and a high daily temperature range (16–30 Celsius; Figure S3; http://clima1.cptec.inpe.br/evolucao/pt) with an average of 27°C, thereby hindering the animals' ability to have progeny earlier. The heritability estimate (0.24 ± 0.04) while incorporating marker information was higher than the traditional pedigree-based estimate reported by Grossi et al. (2008) and Mercadante, Lôbo, and Oliveira (2000) with values equal to 0.14 and 0.09, respectively. AFC heritability estimates are usually low as AFC is a sex-limited trait and heavily influenced by the environment. AFC heritabilities normally range from 0.05 to 0.22 (Grossi et al., 2008). However, AFC can by definition be dissected into three traits: age at the onset of puberty, time from the onset of puberty until conception and gestation length. In our case, age at the onset of puberty is strongly influenced

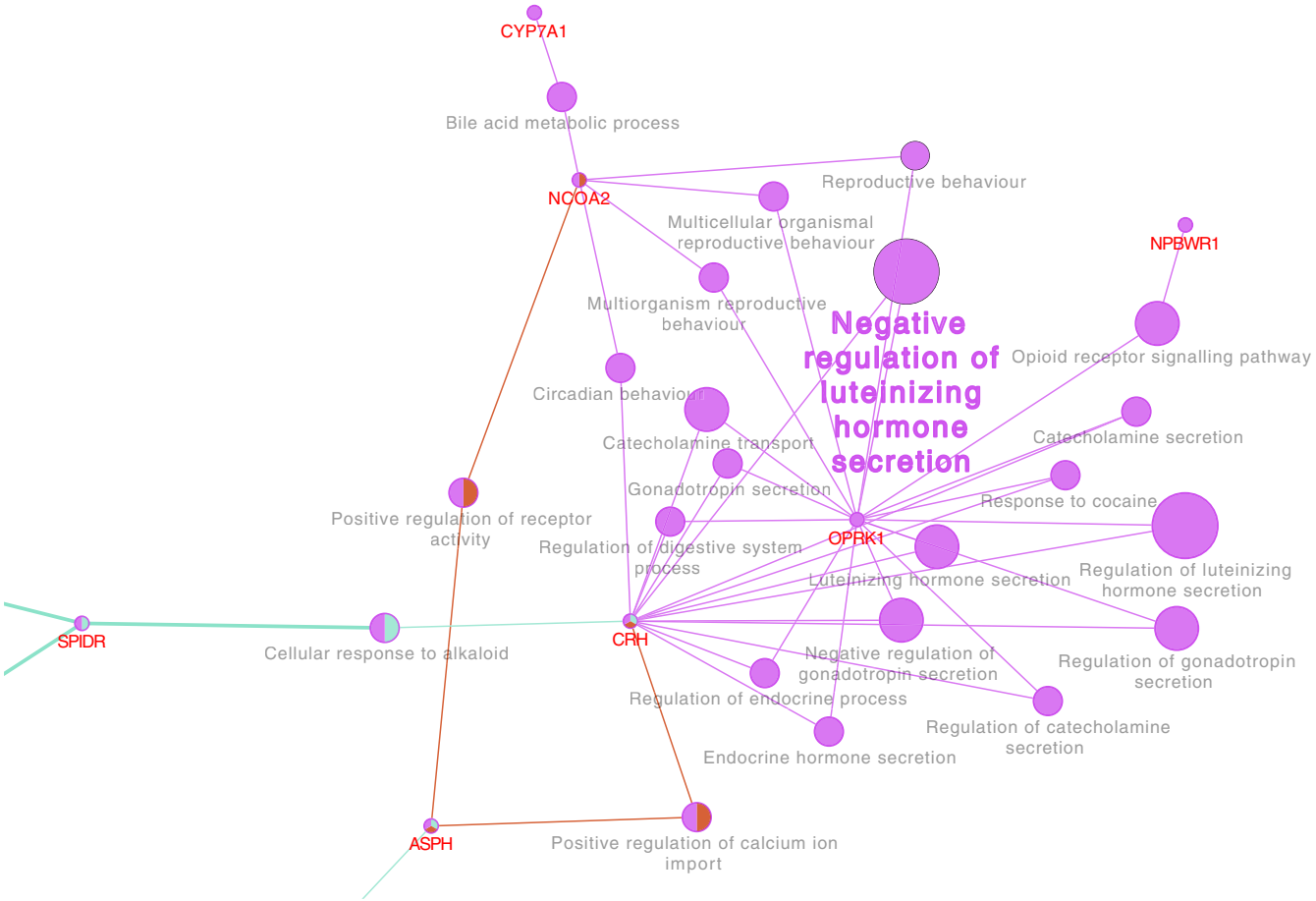

FIGURE 3 Main functional networks with biological process terms and genes (in red) as nodes. The node size represents the term enrichment significance from ClueGO. The most enriched term related age at first calving is shown in bold. [Colour figure can be viewed at wileyonlinelibrary.com]

by the body weight development, which was often reported to be highly heritable (Grossi et al., 2008). It might play a greater role in our study and therefore partially explains the relatively high heritability estimate when compared to other studies and also other fertility traits.

Although genome mapping analysis detected potential chromosome regions explaining the maximum 3% of the genetic variance, this percentage could be considered a relevant proportion, as previously reported in cattle (Utsunomiya et al., 2014). The evidence is that AFC inheritance is complex as it is controlled by multiple loci across the genome. The potential candidate genes that were selected in this study might only describe a small piece of the molecular mechanisms underlying AFC, especially due to the fact that these cows were not bred early (14 or 18 months of age). In addition, if we consider all windows explaining more than 0.40% on BTA14, that is, by summing the genetic variance explained by SNP windows on BTA14, it amounts to nearly 7%. On the other hand, by considering all relevant windows, this value increased to 14.52% (Table 1).

A total of 123 protein-coding genes were observed from the 18 SNP windows, and a biological process network (Figure 3) was built; genes such as *PLAG1*, *NPBWR1*, *OPRK1*, *CRH* and *NCOA2* were highlighted as they showed an important biological process for AFC, such as negative regulation of luteinizing hormone secretion. These four genes are located on BTA14. The *NPBWR1* gene is a neuropeptide W (NPW)/neuropeptide B (NPB) receptor-1 endogenous ligand for the orphan G protein-coupled receptors *GPR7*, which is related to the regulation of feeding and energy metabolism. The endogenous NPW may play a regulatory role in the organization of neuroendocrine signals based on the hypothesis that NPW does not act as a true releasing or inhibiting factor. According to Baker, Cardinal, Bober, Taylor, and Samson (2003), NPW acts in the brain to control prolactin, corticosterone and growth hormone release. However, Fang et al. (2015) reported NPW immunoreactivity levels in sow reproductive systems as intense in primordial follicles, moderate in the uterus and very weak in oocytes of the primary follicle.

The *NPBWR1* gene is located in the adjacent window region that explains the largest proportion (3.09%) of AFC genetic variance mapped on BTA14, peaking between 22.7 and 23.7 Mb (Figure S1). Similarly, the *OPRK1* gene was

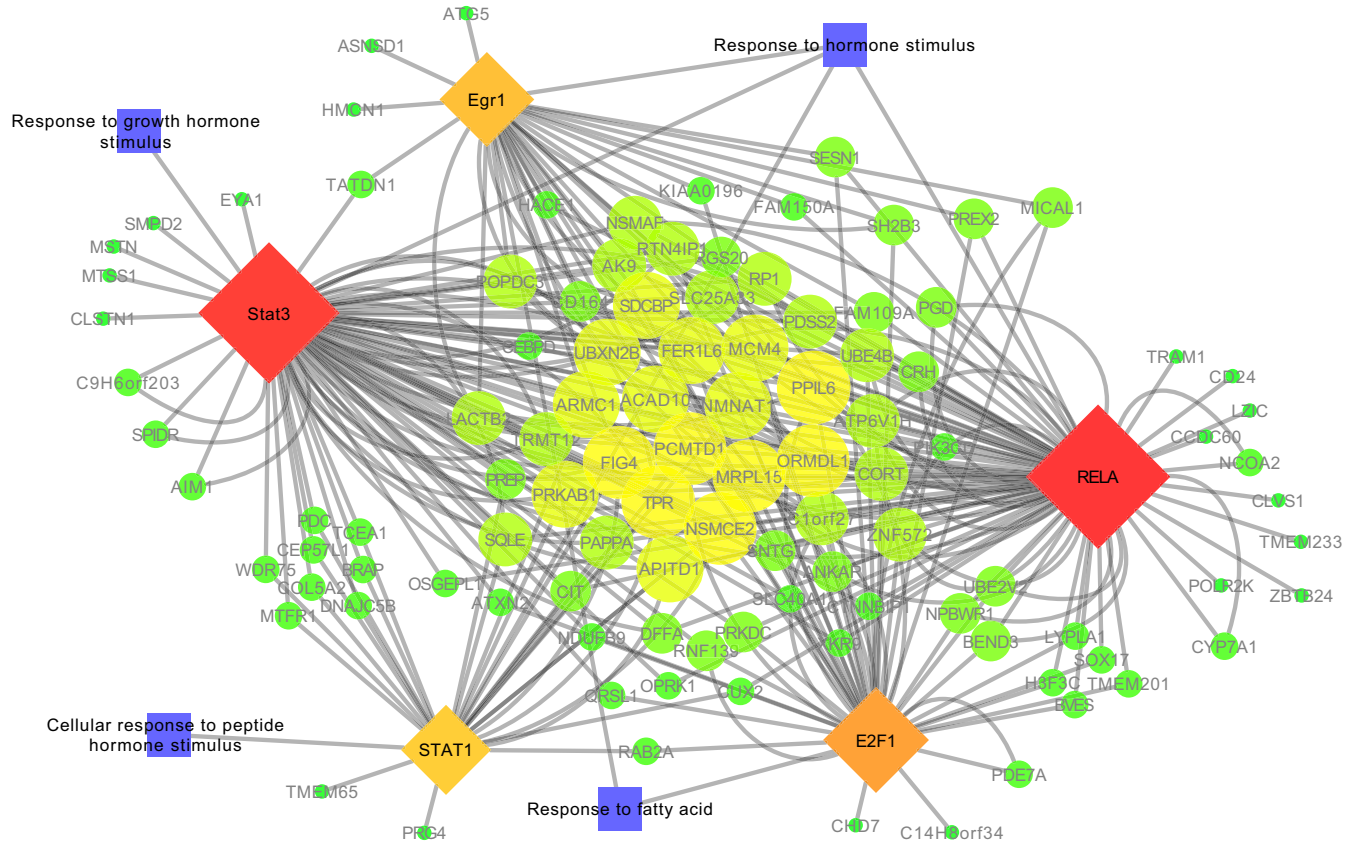

FIGURE 4 Age at first calving (AFC) gene-transcription factor network. Five transcription factors associated with genes involved in AFC: STAT1, STAT3, RELA, E2F1 and EGR1 (diamond nodes), with in silico validated target genes (circle nodes). Their colour scale and size correspond to network analyses (Cytoscape) scores, where red and bigger nodes represent higher edge densities, while green and smaller nodes represent lower edge densities. Blue nodes are the transcription factor-related biological processes. [Colour figure can be viewed at wileyonlinelibrary.com]

also mapped on BTA14. *OPKR1* is an opioid receptor kappa 1 mapped by Fortes et al. (2012) and reported as a potential candidate for Brahman cattle puberty.

The corticotropin-releasing hormone (CRH) gene was identified in another relevant window mapped between 31.25 and 32.25 Mb (Table 1) and was reported to be involved in ovarian steroidogenesis regulation and follicular maturation, ovulation and luteolysis (Kiapekou, Zapanti, Mastorakos, & Loutradis, 2010). Another candidate gene identified on BTA14 was the nuclear receptor coactivator (*NCOA2*) which acts as a transcription factor in the hypothalamus (Fortes et al., 2011). The 1-Mb window containing this protein-coding gene explained around 0.50% of the genetic variance for AFC and was mapped on BTA14 (35.96–36.95 Mb; Table 1). In a study on Brahman cattle, Fortes et al. (2011) reported that *NCOA2* seems to play a key role in the development of puberty by acting as a transcription factor for multiple genes affecting the onset of puberty. Camargo et al. (2015) partially characterized *NCOA2* associated with polymorphisms that were found with reproductive traits in Nellore cattle. These authors have reported significant SNPs in the *NCOA2* gene for early pregnancy probability, days to first calving and AFC in Nellore females.

Biological process analyses from these TF pointed out important reproductive mechanisms (e.g., cellular response to peptide hormone stimulus and response to hormone stimulus) with some literature evidence regarding effects on female puberty, as observed in Table S2. Based on these key TFs, we were able to build a gene-TF network highlighting genes sharing roles with AFC (e.g., *PAPPA*, *PREP*, *PCMTD1*, *FER1L6 NMNAT*, *TPR* and *ACAD10*).

The pregnancy-associated plasma protein-A (*PAPPA*) gene was mapped in an adjacent window between 106.27 and 107.27 Mb on BTA8 and explained 0.52% of AFC genetic variance. This gene is cited to compromise ovarian steroidogenesis and female fertility in mice (Nyegaard et al., 2010). Prolyl endopeptidase (*PREP*) was identified on BTA9 with 0.60% of genetic variance explained. This gene is a member of the serine peptidase group that is widely conserved through evolution. PREP activity has been detected in all organs and tissues with many different types of activities such as maturation and degradation of peptide hormones and neuropeptides. In terms of AFC,

PREP is also related to male and female reproduction-associated processes; gonad function and gamete physiology are some of the roles of *PREP* that were previously described in the literature (Kimura, Matsui, & Takahashi, 2002). These results suggested that PREP might be an important gene influencing AFC, mainly in terms of sire selection.

The protein-L-isoaspartate (D-Aspartate) O-methyltransferase domain containing 1 (PCMTD1) was reported to be expressed in the endometrium of high- and low-fertility heifers during the mid-luteal phase of the oestrous cycle gene in cross-bred cattle (Killeen et al., 2014) and has been cited to be associated with Brahman cattle puberty (Fortes et al., 2012). This gene was mapped on BTA14 in a window region that explained the largest portion (3.09%) of genetic variance located between 22.7 and 23.7 Mb; it was also one of the most highlighted genes in the gene-TF networks and appeared to be one of the most noticeable candidate genes in our study.

Four genes (*FER1L6*, *TPR*, *NMNAT1,* and *ACAD10*) were mapped, respectively, on BTA14, BTA16, BTA16 and BTA17 and are candidate genes with some puberty involvement. The adjacent windows in which these genes are included explained, respectively, 0.41%, 0.47%, 0.85% and 0.55% of AFC genetic variance (Table 1). *FER1L6* is related to folliculogenesis (Stigliani, Anserini, & Nicoletti, 2013), *TPR* to oestrus and early pregnancy in pigs (Goossens & van den Berg, 1979) and *NMNAT1* to female reproductive efficiency in cattle (Khatkar, Randhawa, & Raadsma, 2014), whereas *ACAD10* is involved in fatty acid oxidation during oocyte maturation in mice (Dunning, Anastasi, Zhang, Russell, & Robker, 2014).

Previous studies have found peaks on BTA14 in a similar region, reinforcing this autosome as a chromosome candidate for reproductive traits in cattle. Fortes et al. (2012) identified a large number of SNPs associated with puberty in Brahman cattle. Karim et al. (2011) argued that *PLAG1* was the relevant gene underlying this region on BTA14 and affecting bovine stature; a secondary effect can be related to age at puberty, and therefore to AFC, due to the high genetic correlation between these traits.

Despite the fact we found two SNP markers (BovineHD1400006722 and BovineHD1400006725) inside the gene *PLAG1* (Table S1), we did not confirm this gene as a potential candidate in the present population. Although we reported two close peaks in the region of BTA14 close to *PLAG1*, this region has shown a low LD ($r^2 < .1$; Figures S1 and S4) with the main peak, where the *PCMTD1*, *NPBWR1* and *OPKR1* genes were mapped. The largest peak on BTA14 (Figure 2) segregated into two high LD peaks, located at 22.7–23.7 and 27.8–28.8 Mb. The *PLAG1* gene is located in between (25 Mb) and in low LD ($r^2 < .1$) with both peaks (Table 1; Figure S1), thereby suggesting a segregation of events. Our results suggest that the *PCMTD1*, *NPBWR1* and *OPKR1* genes are strong candidates for AFC on BTA14. Future studies with greater numbers of genotyped and phenotyped animals may confirm these as major genes for AFC in not challenged Nellore cows.

In this study, we identified a high LD within, but low LD between peaks on the same chromosome (Figure S1). Genetic variants within peaks, marked by the associated SNP windows, must have biological functions that hinder the localization of true QTLs related to AFC. Another point to be highlighted is that different genes could be pronounced as potential candidates depending upon the environmental gradient, which may suggest a SNP by environment interaction affecting AFC. In a recently paper, Mota et al. (2017) reported the presence of SNP by environment interactions for tick resistance in cattle affecting different tick burden levels. Therefore, models that incorporate the genotype by environment interaction should be performed.

5 | CONCLUSION

This study contributes important information on AFC genetic variance and the dissection of molecular mechanisms regulating Nellore cattle fertility. The identified potential candidate genes (e.g., *PAPPA, PREP*, *FER1L6*, *TPR*, *NMNAT1*, *ACAD10, PCMTD1, CRH*, *OPKR1*, *NPBWR1* and *NCOA2*) and related gene-TF networks have biological roles that are strongly associated with female reproduction features.

ACKNOWLEDGEMENTS

The authors thank CNPq, FAPEMIG, CAPES (project number 71/2013) and INCT—Ciência Animal for financial support. The Derribadinha farm for provided the data and ABCZ, The University of Queensland and QAAFI for theoretical and computational support.

ORCID

R.R. Mota http://orcid.org/0000-0003-0566-3236
S.E.F. Guimarães http://orcid.org/0000-0003-3704-8131

REFERENCES

Aguilar, I., Misztal, I., Johnson, D., Legarra, A., Tsuruta, S., & Lawlor, T. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science*, *93*(2), 743–752.

Baker, J. R., Cardinal, K., Bober, C., Taylor, M. M., & Samson, W. K. (2003). Neuropeptide W acts in brain to control prolactin,

corticosterone, and growth hormone release. *Endocrinology*, *144* (7), 2816–2821.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., & Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091–1093.

Camargo, G. M., Costa, R. B., Lucia, G., Regitano, L. C., Baldi, F., & Tonhati, H. (2015). Polymorphisms in TOX and NCOA2 genes and their associations with reproductive traits in cattle. *Reproduction, Fertility and Development*, *27*(3), 523–528.

Dunning, K. R., Anastasi, M. R., Zhang, V. J., Russell, D. L., & Robker, R. L. (2014). Regulation of fatty acid oxidation in mouse cumulus-oocyte complexes during maturation and modulation by PPAR agonists. *PLoS One*, *9*(2), e87327.

Fang, R., Su, J., Zheng, L., Jin, M., Hou, Y., Ma, Z., & Ahmed, E. (2015). Cloning and distribution of neuropeptide W and its receptors in pigs. *Research in Veterinary Science*, *101*, 106–116.

Fortes, M., Lehnert, S., Bolormaa, S., Reich, C., Fordyce, G., Corbet, N., & Reverter, A. (2012). Finding genes for economically important traits: Brahman cattle puberty. *Animal Production Science*, *52* (3), 143–150.

Fortes, M. R., Li, Y., Collis, E., Zhang, Y., & Hawken, R. J. (2013). The IGF1 pathway genes and their association with age of puberty in cattle. *Animal Genetics*, *44*(1), 91–95.

Fortes, M. R., Reverter, A., Nagaraj, S. H., Zhang, Y., Jonsson, N. N., Barris, W., & Hawken, R. J. (2011). A single nucleotide polymorphism-derived regulatory gene network underlying puberty in 2 tropical breeds of beef cattle. *Journal of Animal Science*, *89*(6), 1669–1683.

Goossens, J., & van den Berg, T. P. (1979). Possible aspects of normal and unphysiological commercial conditions on the estrus and early pregnancy in pigs. *Veterinary Quarterly*, *1*(2), 109–114.

Grossi, D., Frizzas, O., Paz, C., Bezerra, L., Lôbo, R., Oliveira, J., & Munari, D. (2008). Genetic associations between accumulated productivity, and reproductive and growth traits in Nelore cattle. *Livestock Science*, *117*(2), 139–146.

Hanna, L. L. H., & Riley, D. G. (2014). Mapping genomic markers to closest feature using the R package Map2NCBI. *Livestock Science*, *162*, 59–65.

Karim, L., Takeda, H., Lin, L., Druet, T., Arias, J. A., Baurain, D., & Grisart, B. (2011). Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nature Genetics*, *43*(5), 405–413.

Khatkar, M., Randhawa, I., & Raadsma, H. (2014). Meta-assembly of genomic regions and variants associated with female reproductive efficiency in cattle. *Livestock Science*, *166*, 144–157.

Kiapekou, E., Zapanti, E., Mastorakos, G., & Loutradis, D. (2010). Update on the role of ovarian corticotropin-releasing hormone. *Annals of the New York Academy of Sciences*, *1205*(1), 225–229.

Killeen, A. P., Morris, D. G., Kenny, D. A., Mullen, M. P., Diskin, M. G., & Waters, S. M. (2014). Global gene expression in endometrium of high and low fertility heifers during the mid-luteal phase of the estrous cycle. *BMC Genomics*, *15*(1), 234.

Kimura, A., Matsui, H., & Takahashi, T. (2002). Expression and localization of prolyl oligopeptidase in mouse testis and its possible involvement in sperm motility. *Zoological Science*, *19*(1), 93–102.

Mercadante, M. E. Z., Lôbo, R. B., & Oliveira, H. D. (2000). Estimativas de (co) variâncias entre características de reprodução e de crescimento em fêmeas de um rebanho Nelore. *Revista Brasileira de Zootecnia*, *29*(4), 997–1004.

Mota, R. R., Silva, F. F., Lopes, P. S., Tempelman, R. J., Sollero, B. P., Aguilar, I., & Cardoso, F. F. (2017). Analyses of reaction norms reveal new chromosome regions associated with tick resistance in cattle. *Animal*. https://doi.org/10.1017/S1751731117001562

Nyegaard, M., Overgaard, M. T., Su, Y.-Q., Hamilton, A. E., Kwintkiewicz, J., Hsieh, M., & Giudice, L. C. (2010). Lack of Functional Pregnancy-Associated Plasma Protein-A (PAPPA) compromises mouse ovarian steroidogenesis and female fertility 1. *Biology of Reproduction*, *82*(6), 1129–1138.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research*, *32* (suppl 1), D91–D94.

Sargolzaei, M., Chesnais, J., & Schenkel, F. (2011). FImpute-An efficient imputation algorithm for dairy cattle populations. *Journal of Dairy Science*, *94*(1), 421.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.

Stigliani, S., Anserini, P., & Nicoletti, A. (2013). Insight into the genomics of premature ovarian failure. *Molecular Genetics & Genomic Medicine*, *7*(3), 1–7.

Touzet, H., & Varré, J.-S. (2007). Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms for Molecular Biology*, *2*(1), 15.

Utsunomiya, Y. T., Carmo, A. S., Neves, H. H., Carvalheiro, R., Matos, M. C., Zavarez, L. B., & Porto-Neto, L. R. (2014). Genome-wide mapping of loci explaining variance in scrotal circumference in Nellore cattle. *PLoS One*, *9*(2), e88561.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., & Muir, W. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genetics Research*, *94*(02), 73–83.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., & Sonstegard, T. S. (2009). A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology*, *10*(4), R42.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Mota RR, Guimarães SEF, Fortes MRS, et al. Genome-wide association study and annotating candidate gene networks affecting age at first calving in nellore cattle. *J Anim Breed Genet*. 2017;134:484–492. https://doi.org/10.1111/jbg.12299