



# Female flowers first: QTL mapping in eastern black walnut (*Juglans nigra* L.) identifies a dominant locus for heterodichogamy syntenic with that in Persian walnut (*J. regia* L.)

Warren Chatwin<sup>1</sup> · Devin Shirley<sup>2</sup> · Jacqueline Lopez<sup>3</sup> · Joseph Sarro<sup>4</sup> · John Carlson<sup>5</sup> · Alison Devault<sup>6</sup> · Mike Pfreder<sup>4,7</sup> · Ronald Revord<sup>8</sup> · Mark Coggeshall<sup>8</sup> · Jeanne Romero-Severson<sup>7</sup>

Received: 8 July 2022 / Revised: 21 November 2022 / Accepted: 30 November 2022 / Published online: 17 December 2022  
© The Author(s) 2022

## Abstract

Eastern black walnut (*Juglans nigra* L.), one of the most valuable timber and veneer trees in North America, provides nut shells with unique industrial uses and nut kernels with distinctive culinary attributes. A mature F<sub>1</sub> full-sib progeny orchard of 248 individuals from the cross of two eastern black walnut cultivars provides a long-term resource for discovering genetic mechanisms controlling life history, quality traits, and stress resistance. The genetic linkage map, constructed with 356 single nucleotide polymorphism (SNP) markers and 62 expressed sequence tag simple sequence repeats (EST-SSRs), is 1645.7 cM in length, distributed across the expected 16 linkage groups. In this first application of QTL mapping in *J. nigra*, we report QTL for budbreak, peak pistillate bloom, peak staminate bloom, and heterodichogamy. A dominant major QTL for heterodichogamy is reported, the sequence for which is syntenic with the heterodichogamy QTL on chromosome 11 of Persian walnut (*J. regia* L.). The mapping population parents are both protogynous, and segregation suggests a Mendelian component, with a 3:1-like inheritance pattern from heterozygous parents. Mapping the sequenced EST-SSR markers to the *J. regia* “Chandler” V2.0 genome sequence revealed evidence for collinearity and structural changes on two of the sixteen chromosomes. The inclusion of sequenced EST-SSR markers enables the direct comparison of this and subsequent *J. nigra* maps and other *Juglandaceae* genetic maps. This investigation initiates long-term QTL detection studies for quality and stress resistance traits in black walnut.

**Keywords** *Juglans* · Eastern black walnut · Persian walnut · EST-SSR · Heterodichogamy · QTL mapping

Communicated by D. Chagné

✉ Ronald Revord  
r.revord@missouri.edu

- <sup>1</sup> United States Department of Agriculture – Agricultural Research Service, Crop Germplasm Research Unit, College Station, TX, USA
- <sup>2</sup> Utah Division of Wildlife Resources, Bountiful, UT, USA
- <sup>3</sup> Genomics & Bioinformatics Core Facility, University of Notre Dame, Notre Dame, IN, USA
- <sup>4</sup> Genetics Bioinformatics Service Center, Stanford University School of Medicine, Stanford, CA, USA
- <sup>5</sup> The Pennsylvania State University, University Park, PA, USA
- <sup>6</sup> Daicel Arbor Biosciences, Ann Arbor, MI, USA
- <sup>7</sup> Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, USA
- <sup>8</sup> School of Natural Resources, University of Missouri Center for Agroforestry, Columbia, MO, USA

## Introduction

The economic and ecological value of many long-lived forest trees is far higher than the cost of determining the genetic underpinnings of the forest tree characteristics essential for ecosystem services and the specific traits humans value. This disparity is especially evident in walnuts, the *Juglans* genus. Persian walnut (*Juglans regia* L.), native to Central Asia, is the most commonly cultivated nut tree species in the world (Martínez-García et al. 2016). The nut crop in California alone was valued at more than 1.2 billion in 2019 (NASS 2020). Eastern black walnut (*J. nigra* L.), native to North America, is a highly valued timber, veneer, and nut-producing species with standing growing stock value in North America estimated at over half a trillion US dollars (Daniels et al. 2016). The high value of *J. regia* nut crop and the concentration of production in orchards in California has spurred recent

advances in the detection of the genetic mechanisms for quality and resistance traits (Bernard et al. 2020).

In contrast, most of the highest value *J. nigra* timber is taken from small private holdings widely distributed across the central USA. Timber, nuts, and nutshells, which provide a grit with unique metal polishing properties, are harvested primarily from wild trees (Coggeshall 2011). However, rapid advances in genomics technology, the steeply declining cost of using the technology, and the maturation of *J. nigra* pedigreed populations now enable us to initiate QTL mapping and compare the genetic architecture of *J. nigra* with that of *J. regia*. Although interspecific hybrids are rare, synteny and high collinearity between these two species would suggest that the location of major genes and the genetic architecture of complex traits in one species may provide insight into the same features in the other (Pollegioni et al. 2013). The information already gained on the genetic mechanisms that confer high-quality nut traits in *J. regia* could be directly applicable for nut quality in *J. nigra*.

*J. regia* and *J. nigra* have markedly different recent histories. *J. regia*, native in Central Asia, has experienced a long period of domestication and dissemination outward along the Silk Road over thousands of years (Pollegioni et al. 2015; Zeven and Zhukovsky 1975). Recent evidence suggests that human-mediated movement of *J. regia* from Asia to the Balkans and Europe began in the Bronze Age (Pollegioni et al. 2017; Roor et al. 2017). During this time, farmers contributed to genetic improvement through seedling selection and clonal propagation (McGranahan and Leslie 2012). Organized breeding began over 70 years ago, starting in the twentieth century (Bernard et al. 2018). Early efforts combined high nut quality, lateral bearing, precocity, and early harvest time (Bernard et al. 2017; Vahdati et al. 2019). In recent decades, the genetic improvement of *J. regia* has grown to represent many regions globally, resulting in many new cultivars, selections, and an expanded emphasis on abiotic and biotic stress tolerance (Bernard et al. 2017; Vahdati et al. 2019). In contrast, *J. nigra* remained undomesticated until the twentieth century, when growers and tree breeders selected and propagated improved selections of *J. nigra* for both timber and nut production from on-farm and naturally regenerated stands (MacDaniels 1933; Talbert 1942). The first improved clonal releases of *J. nigra* for high timber quality have recently transitioned from progeny tests (assessing timber-form, growth, insect, and disease resistance) to improved clonal orchards with test sites across multiple environments (McKenna and Coggeshall 2018). Nearly 700 *J. nigra* cultivars were named during the twentieth century based on nut production and quality observations, most of which were chance wild seedlings selected by growers and hobbyists (Michler et al. 2007; Reid et al. 2004).

In a long-term effort to produce black walnut cultivars for nut production, the University of Missouri Center for Agroforestry (UMCA) has curated and maintained a collection

of over 65 high kernel percentage cultivars of *J. nigra* since 1996 (Coggeshall 2011). Cultivars from this collection (with kernel percentages in the upper 20 s to lower 30 s) were used to initiate the first organized genetic improvement program for nut/kernel producing *J. nigra* in 2001. The phenotypes measured across many years in this collection include leafing date, protandry, nut ripening time, percent kernel, inter-annual yield, and anthracnose tolerance (Meier et al. 2022). The parents of the *J. nigra* mapping population in this study were chosen from this collection.

Efforts to discover marker-trait associations in *Juglans* were previously limited to *J. regia*. Three initial studies utilizing different bi-parental mapping populations found QTL for anthracnose resistance (Zhu et al. 2015), lateral bearing, harvest date, nut quality traits (Aradhya et al. 2019), and leafing time (Kefayati et al. 2019). More recently, a series of QTL mapping and genome-wide association studies (GWAS) have significantly expanded the volume of traits with known QTLs and associated DNA markers, including nut suture strength (Sideli et al. 2020b), pellicle pigment/color (Marrano et al. 2019b; Sideli, et al. 2020a), many phenological traits, and lateral bearing (Bernard et al. 2020; Marrano et al. 2019b). The Axiom™ *J. regia* 700 K SNP genotyping array (Marrano et al. 2019a) was used for genotyping in these studies, providing an excellent basis for comparative mapping within *J. regia*.

The SNP, genomic SSR, and SLAF markers used in these *J. regia* QTL mapping studies are likely to have limited transferability to the Rhysocaryon, a walnut section native to North America. Reduced representation sequencing (RRS), widely used in forest trees to generate genetic maps and perform GWAS, produces genetic markers with short sequence lengths (50–100 bp) that have relatively low transferability between populations of the same or related species (Grattapaglia et al. 2011; Lepoittevin et al. 2015; Montanari et al. 2013). The *J. regia* “Chandler” v2.0 genome, with a chromosome-scale resolution, was recently published (Marrano et al. 2020); and the draft assemblies for *J. sigillata*, *J. nigra*, *J. microcarpa*, *J. hindsii*, and *J. cathayensis* were published in 2018 (Stevens et al. 2018). These whole-genome sequencing projects enable the direct assessment of chromosome architecture across species for selected individuals but are costly to employ in QTL mapping or GWAS studies. Expressed sequence tag single sequence repeats (EST-SSR) markers, detected in expressed genes containing repeated sequences, are robust and polymorphic DNA markers transferable across all genus members and closely related genera. EST-SSR are exceptionally useful for genetic mapping, QTL detection, complex interspecific ancestry assessment in large samples, and initial comparative genomics studies (Bradbury et al. 2013; Leonarduzzi et al. 2016; McCleary et al. 2013; Santos et al. 2017; Torokeldiev et al. 2018).

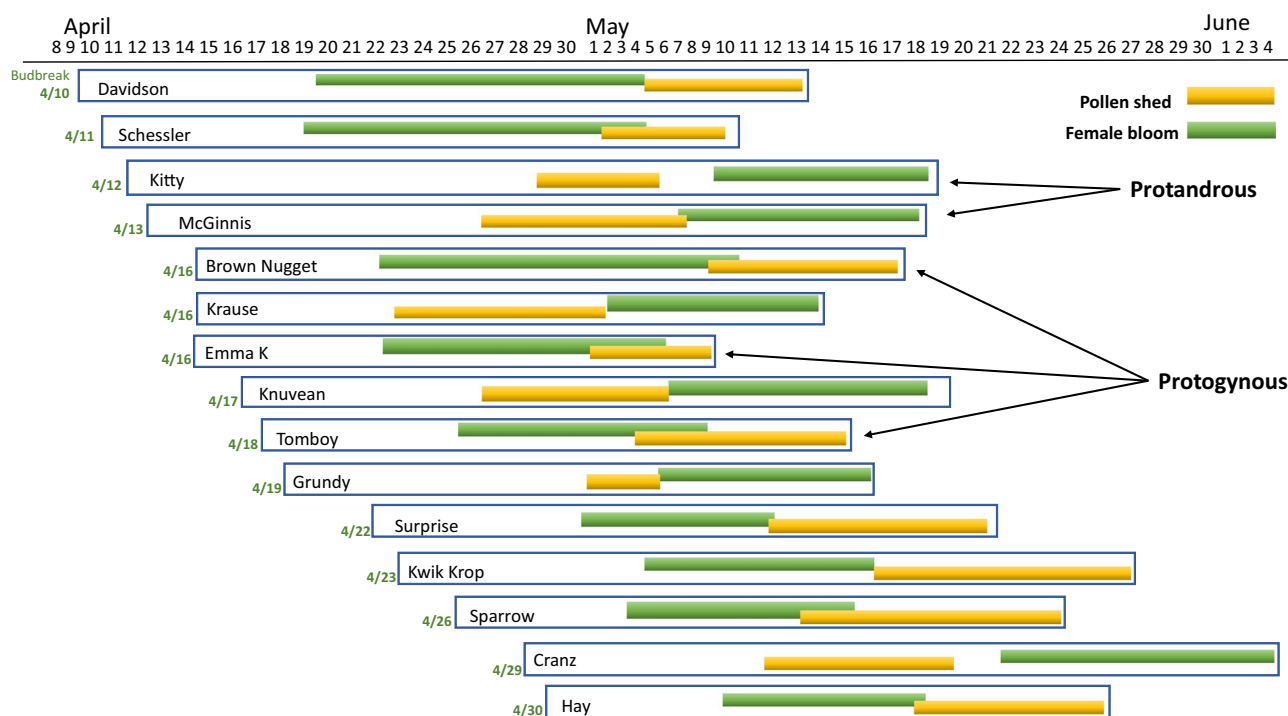
An understanding of the genetic mechanisms controlling floral synchronization is essential for genetic gain in breeding programs, forest resource management, and the study of angiosperm evolution, including speciation. Monecious species exhibit many variants on the timing of pistillate and staminate bloom, including dichogamy, in which the two sexes flower at different times in the same year (Braun and Gottsberger 2011; Cruden 1988; Lloyd and Webb 1986; Renner 2001). Floral synchronization in the monecious Juglandaceae is a heterodichogamous system (Endress 2020). Heterodichogamy is characterized by two temporal genetic morphs in the same species (Fukuhara and Tokumaru 2014). In the protogynous morph, pistillate flowers mature before the staminate flowers. In the protandrous morph, staminate flowers mature before pistillate flowers. The first gene identified in pecan (*Carya illinoensis*) was a single Mendelian factor determining the dichogamy morph, with protogyny dominant over protandry. The observation of segregating populations in *J. regia*, *J. ailantifolia*, *J. hindsii*, and other studies in *C. illinoensis* also indicate a single locus, with protogyny dominant to protandry (Gleeson 1982; Kimura et al. 2012; Kimura et al. 2003; Thompson and Romberg 1985). A single QTL locus for heterodichogamy maps to chromosome 11 in *J. regia* (Bernard et al. 2020) and chromosome 4 in *C. illinoensis* (Bentley et al. 2019).

This investigation combines the high information content of sequenced EST-SSR markers and RRS-generated SNP markers to construct the first genetic map for *J. nigra*. The mapping population consisted of 248 same age full-sibling individuals of “Sparrow” × “Schessler.” Both cultivars are protogynous. We report a QTL analysis of the first life history traits (heterodichogamy, budbreak, pistillate bloom, and staminate bloom), including a major locus for heterodichogamy that is syntenic to that discovered in *J. regia*. Finally, our use of sequenced EST-SSR in the genetic map produced evidence of collinearity between *J. nigra* and *J. regia* and multiple rearrangements.

## Results

### Variation for phenological traits

The 248 individuals of the “Sparrow” × “Schessler” mapping population were evaluated for budbreak in 2016, 2017, 2020, and 2021 and peak pistillate bloom, peak staminate bloom, and heterodichogamy in 2020 and 2021. Average phenological time of “Sparrow” and “Schessler” is displayed in Fig. 1. While all mapping population individuals were evaluated for budbreak in the listed years, not all trees were



**Fig. 1** Budbreak and bloom times for nine protogynous and six protandrous *J. nigra* cultivars in New Franklin, MO. The distribution across years (2002–2006) for the interval in days between the 50%

bloom dates of pistillate and staminate flowers and the phenology of blooming dates for 11 black walnut cultivars, including the parents of the mapping population “Sparrow” and “Schessler”

flowering in 2020 and 2021, and thus, the number of individuals used for these analyses were less than 248 and based on the available phenotypic data (Table 1).

Budbreak, peak pistillate bloom, and peak staminate bloom exhibited a high range but moderately high standard deviations, with basic statistics summarized in Table 2. Budbreak populations means varied from a low of  $112.05 \pm 4.60$  Julian days in 2017 to a high of  $120.73 \pm 2.36$  in 2021. The largest range of budbreak occurred in 2017 from 100 to 128 Julian days. Conversely, a late, cold 2021 spring delayed the start of budbreak to 111 Julian day, with a compressed range to 130. Peak pistillate and staminate blooms were relatively consistent between 2020 ( $136.15 \pm 4.04$  and  $141.51 \pm 5.14$ , respectively) and 2021 ( $134.69 \pm 3.98$  and  $139.79 \pm 4.95$ , respectively) (Table 2), with mean peak pistillate bloom occurring about five Julian days earlier. Interestingly, the standardization of budbreak was markedly lower than that of the flowering time traits, although not in all years (i.e., 2017). Peak pistillate bloom ranged from 124.5 Julian day to 146.5 on average, while the range of average peak staminate bloom was similar.

Heterodichogamy displayed the least variation with most individuals of the protogynous morph (70 out of 97 in 2020 and 83 or 112 in 2021), about three-fourths of the population. The remaining individuals represented the homogamous morph (17 and 25 individuals in 2020 and 2021, respectively) and the protandrous morph (10 and 4 individuals in 2020 and 2021, respectively). In the protogynous individuals, the peak bloom dates for pistillate flowers ranged from 125 Julian day to 141. The interval between peak pistillate bloom and the peak staminate bloom ranged from 2 to 19 days. In the protandrous individuals, the peak bloom dates for the staminate flowers ranged from 126 Julian day to 140, and the interval between peak staminate bloom and the peak pistillate bloom ranged from 1 to 12 days (Fig. 2). The pistillate flowers of protogynous progeny reached peak bloom an average of 4 days earlier than the pistillate flowers of protandrous (and homogamous) progeny ( $p < 0.001$ ). The staminate flowers of protogynous progeny reached peak bloom an average of 8 days later than the staminate flowers

**Table 1** The number of individuals per phenotypic data class per year and the combined unique incorporated into best linear unbiased predictions

|                           | Budbreak | Peak pistillate bloom | Peak staminate bloom | Heterodichogamy |
|---------------------------|----------|-----------------------|----------------------|-----------------|
| Combined unique genotypes | 248      | 201                   | 152                  | 130             |
| 2020                      | 248      | 151                   | 132                  | 97              |
| 2021                      | 248      | 189                   | 129                  | 112             |

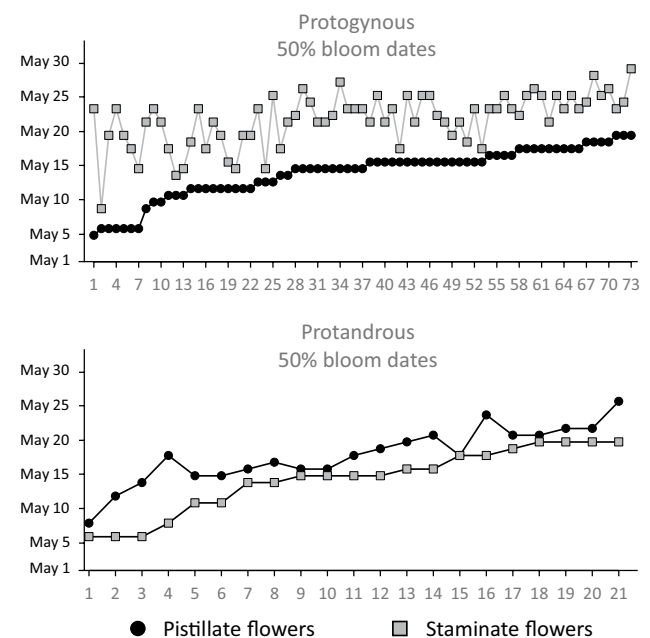
**Table 2** Summary statistics of phenological traits

| Trait                 | Year | Mean <sup>a</sup> | SD <sup>b</sup> | Range   |
|-----------------------|------|-------------------|-----------------|---------|
| Budbreak              | 2016 | 113.62            | 3.24            | 103–124 |
|                       | 2017 | 112.05            | 4.60            | 100–128 |
|                       | 2020 | 117.81            | 3.24            | 105–124 |
|                       | 2021 | 120.73            | 2.36            | 111–130 |
| Peak pistillate bloom | 2020 | 136.15            | 4.04            | 126–147 |
|                       | 2021 | 134.69            | 3.98            | 123–146 |
| Peak staminate bloom  | 2020 | 141.51            | 5.14            | 127–151 |
|                       | 2021 | 139.79            | 4.95            | 123–150 |
| Heterodichogamy       | 2020 | 6.63              | 1.87            | 1–9     |
|                       | 2021 | 7.01              | 1.63            | 1–9     |

<sup>a</sup>Phenological traits are in Julian days, and heterodichogamy is expressed in the International Plant Genetic Resource Institutions (1994) 1–9 ordinal class

<sup>b</sup>Standard deviation

of protandrous (and homogamous) progeny ( $p < 0.001$ ). The mean interval between pistillate and staminate bloom in protogynous individuals was significantly longer than the mean interval between pistillate and staminate bloom in protandrous individuals (8 days vs. 3.8 days,  $p < 0.001$ , Fig. 2).



**Fig. 2** Bloom date and interval between pistillate and staminate bloom date for protogynous and protandrous individuals in the  $F_1$  mapping population. Bloom dates are recorded as the date of 50% bloom for the first sex to flower and then the date of 50% bloom for the next sex to flower, in the same individual. Protogynous and protandrous individuals shown separately for clarity. The y-axes are identical for both. Pistillate bloom dates are indicated by black circles. Staminate bloom dates are indicated as gray boxes

Average dates of budbreak and bloom dates collected over a 5-year interval (2002–2006) on 15 cultivars (including the mapping parents) did not suggest a close association between the timing of budbreak, the morph type, and the interval between the flowering of the two morphs on the same cultivar (Fig. 1).

## Marker yields and characteristics

### ddRADseq

ddRADseq yielded 27,375 SNP-containing fragments 80–110 bp in length. SNPs located in the first nine bp of the sequence were removed along with sequences containing multiple SNPs. Experience with RRS technologies indicated that SNPs in the first nine bp of a sequenced fragment are much more likely to be errors, while fragments with more than one SNP are likely to be members of clusters of highly repetitive sequences (Konar et al. 2017). The remaining markers were filtered for informative marker configurations in our two outcrossing parents, resulting in 9603 SNPs. Markers in the heterozygous pseudo- $F_1$  configuration (hk x hk), 2226, were less frequent than markers in the pseudo-testcross configuration (3482 lm x ll and 3895 nn x np). The median values for missing data ranged from 35.2% (hk x hk) to 37% for both pseudo-testcross configurations. Individual progeny with more than 90% missing data (26 individuals) and 8585 markers with more than 15% missing data were removed, leaving 1257 markers for mapping.

### EST-SSR

The EST-SSR bait-capture approach generated sequences for 681 markers on 149 individuals. The number of individuals sequenced was a function of funding limitations. The majority of EST-SSR markers (407) met our criterion for missing data (< 50% missing data). Two hundred and sixty-six of these were monomorphic for the SSR repeat number, 93 from *J. nigra* and 181 from *J. regia*. Eight EST-SSR markers from *J. nigra* were homozygous for the SSR repeat number in both parents, but the alleles for each parent were different. This configuration was therefore uninformative in the progeny. After removing heterozygous markers with more than 15% missing data, 118 markers remained for mapping.

## Linkage mapping

The *J. nigra* “Sparrow” x “Schessler” map was constructed using 248  $F_1$  progeny genotyped with ddRADseq, 149 of which were also genotyped using EST-SSR markers. The mapping method employed by JoinMap 4.1 uses the markers informative in both parents and the markers occurring in the pseudo-testcross configuration (informative only in the

male or female parent) to generate one consensus map rather than separate male and female maps. Of the 118 EST-SSR markers and 1257 SNP markers that passed the missing data thresholds, only 79 EST-SSR markers and 590 SNP markers remained after grouping, chi-square analyses, and evaluations of nearest-neighbor fit and stress. Seventeen EST-SSRs and 234 SNPs had sufficient linkage to group, but insufficient linkage for genetic mapping. The final consensus map contains 62 EST-SSR markers and 256 SNP markers with a total length of 1646 cM (Table 3, Online Resource 1). The average marker density was 4.2 cM, with 19 regions having gaps greater than 10 cM. The average gap length was 12.9 cM.

## QTL mapping of budbreak, peak bloom dates, and heterodichogamy

QTL mapping results are summarized in Table 4. Five minor QTL (linkage group (LG)2, LG7, LG9, LG10, and LG11) were discovered for budbreak when analyzing 4-year BLUP data without a covariate. Two QTL were still detected on LG2 and LG11 for budbreak when introducing peak pistillate bloom as a covariate (Table 4). Covariance arises because the pistillate flower emerges from the same bud as the leaves. The QTL on LG7, LG9, and LG10 are likely to be the confounding effects of phenotypic covariance for peak pistillate bloom. The staminate flowers emerge from their own buds; thus, we did not use peak staminate bloom as a covariate. The LG2 QTL was detected at a LOD of 7.2 (0.99 significance threshold = 5.1), explaining 7.4% of the variance and supporting a 95% confidence interval (CI) from 70.1 to 80.3 cM. The LG11 QTL supported a LOD of 6.3 and a 95% CI from 0.0 to 6.0 cM, explaining 6.5% of the variation.

QTL mapping for heterodichogamy detected a significant QTL region spanning 21 cM on LG11. Significant QTL were not detected elsewhere in the genome. The sequences of the four markers spanning the highest peak ( $p < 0.01$ ) 165067\_18, 16715\_92, estJN1060, and 49576\_83, map to a genomic region associated with heterodichogamy in a 170 accession *J. regia* GWAS panel sequenced with the Axiom™ *J. regia* 700 K SNP array (Fig. 3) (Bernard et al. 2020). This result validates our version one genetic map for *J. nigra*, despite the comparatively small mapping population size and number of markers.

In this mapping population, the pistillate flowers of protogynous progeny reached peak bloom an average of 4 days earlier than the pistillate flowers of protandrous progeny ( $p < 0.001$ ) and the staminate flowers of protogynous progeny reached peak bloom an average of 8 days later than the staminate flowers of protandrous progeny ( $p < 0.001$ ); thus, the phenotype heterodichogamy is a phenotypic covariate for peak female and male bloom dates. When using

**Table 3** Statistics for all 16 linkage groups produced in our genetic map. Numbers in parentheses represent expected progeny segregation ratios for those marker configurations

| LGs | LG length (cM) | Avg. marker distance (cM) | No. of markers | No. EST-SSRs | No. SNPs | ab x cd (1:1:1:1) | ef x eg (1:1:1:1) | hk x hk (1:2:1) | lm x ll (1:1) | nn x np (1:1) |
|-----|----------------|---------------------------|----------------|--------------|----------|-------------------|-------------------|-----------------|---------------|---------------|
| 1   | 144.9          | 4.527                     | 33             | 4            | 29       | 0                 | 2                 | 4               | 17            | 10            |
| 2   | 122.1          | 4.070                     | 31             | 9            | 22       | 0                 | 4                 | 4               | 15            | 8             |
| 3   | 74.1           | 2.646                     | 29             | 6            | 23       | 0                 | 0                 | 4               | 13            | 12            |
| 4   | 107.5          | 4.885                     | 23             | 2            | 21       | 0                 | 0                 | 5               | 11            | 7             |
| 5   | 81.7           | 5.444                     | 16             | 2            | 14       | 0                 | 0                 | 6               | 4             | 6             |
| 6   | 115.6          | 5.502                     | 22             | 1            | 21       | 0                 | 1                 | 1               | 6             | 14            |
| 7   | 128.9          | 3.681                     | 36             | 5            | 31       | 1                 | 0                 | 6               | 10            | 19            |
| 8   | 78.2           | 4.113                     | 20             | 2            | 18       | 0                 | 0                 | 4               | 11            | 5             |
| 9   | 91.5           | 3.979                     | 24             | 4            | 20       | 0                 | 2                 | 3               | 12            | 7             |
| 10  | 94.4           | 3.629                     | 27             | 5            | 22       | 0                 | 0                 | 6               | 13            | 8             |
| 11  | 113.2          | 4.352                     | 27             | 5            | 22       | 0                 | 0                 | 7               | 9             | 11            |
| 12  | 99.3           | 4.514                     | 23             | 3            | 20       | 0                 | 0                 | 10              | 8             | 5             |
| 13  | 107.5          | 2.984                     | 37             | 5            | 32       | 0                 | 2                 | 8               | 16            | 11            |
| 14  | 86.6           | 3.462                     | 26             | 3            | 23       | 0                 | 1                 | 8               | 13            | 4             |
| 15  | 81.5           | 5.096                     | 17             | 2            | 15       | 0                 | 0                 | 3               | 6             | 8             |
| 16  | 118.7          | 4.567                     | 27             | 4            | 23       | 1                 | 1                 | 9               | 9             | 7             |

**Table 4** Summary of quantitative trait loci analysis results for budbreak, peak pistillate bloom, peak staminate bloom, and heterodichogamy

| Trait                 | Covariate             | Position | Marker        | LG | LOD | Sig. threshold <sup>a</sup> | 95% CI <sup>b</sup> | P.E.V. <sup>c</sup> |
|-----------------------|-----------------------|----------|---------------|----|-----|-----------------------------|---------------------|---------------------|
| Budbreak              |                       | 71.6     | SNP 53772_16  | 2  | 5.4 | 5.3                         | 60.6–90.1           | 9.5                 |
|                       |                       | 111.8    | SNP 230028_88 | 7  | 4.8 | 4.5                         | 85.4–125.9          | 8.5                 |
|                       |                       | 49.2     | SNP 176789_94 | 9  | 5.4 | 5.3                         | 38.4–61.3           | 9.6                 |
|                       |                       | 92.7     | SNP 281018_28 | 10 | 6.9 | 5.3                         | 84.7–94.3           | 12                  |
|                       |                       | 5.0      | SNP 328893_63 | 11 | 5.3 | 5.3                         | 0.0–21.1            | 9.4                 |
| Budbreak              | Peak pistillate bloom | 71.6     | SNP 53772_16  | 2  | 7.2 | 5.1                         | 70.1–80.3           | 7.4                 |
|                       |                       | 0.0      | SNP 237350_80 | 11 | 6.3 | 5.1                         | 0.0–6.0             | 6.5                 |
| Peak pistillate bloom | Heterodichogamy       | 13.5     | estJN1025     | 11 | 7.1 | 5.3                         | 2.0–22.1            | 18.4                |
| Peak staminate bloom  |                       | 88.9     | SNP 73132_43  | 2  | 5.5 | 4.3                         | 70.1–91.7           | 15.5                |
|                       |                       | 20.1     | estJN1060     | 11 | 18  | 4.1                         | 13.5–24.7           | 42.5                |
| Peak staminate bloom  | Heterodichogamy       | 14.5     | SNP 229102_28 | 11 | 12  | 5.5                         | 0.0–19.1            | 22.6                |
| Heterodichogamy       | -                     | 26.7     | SNP 49576_83  | 11 | 9.5 | 5.4                         | 21.1–40.3           | 28.8                |

<sup>a</sup>Significance threshold is based on the genome-wide LOD value of a 10,000-iteration permutation test that corresponds to a 0.99 probability, with exception to budbreak on LG 7 where a 0.95 probability was used

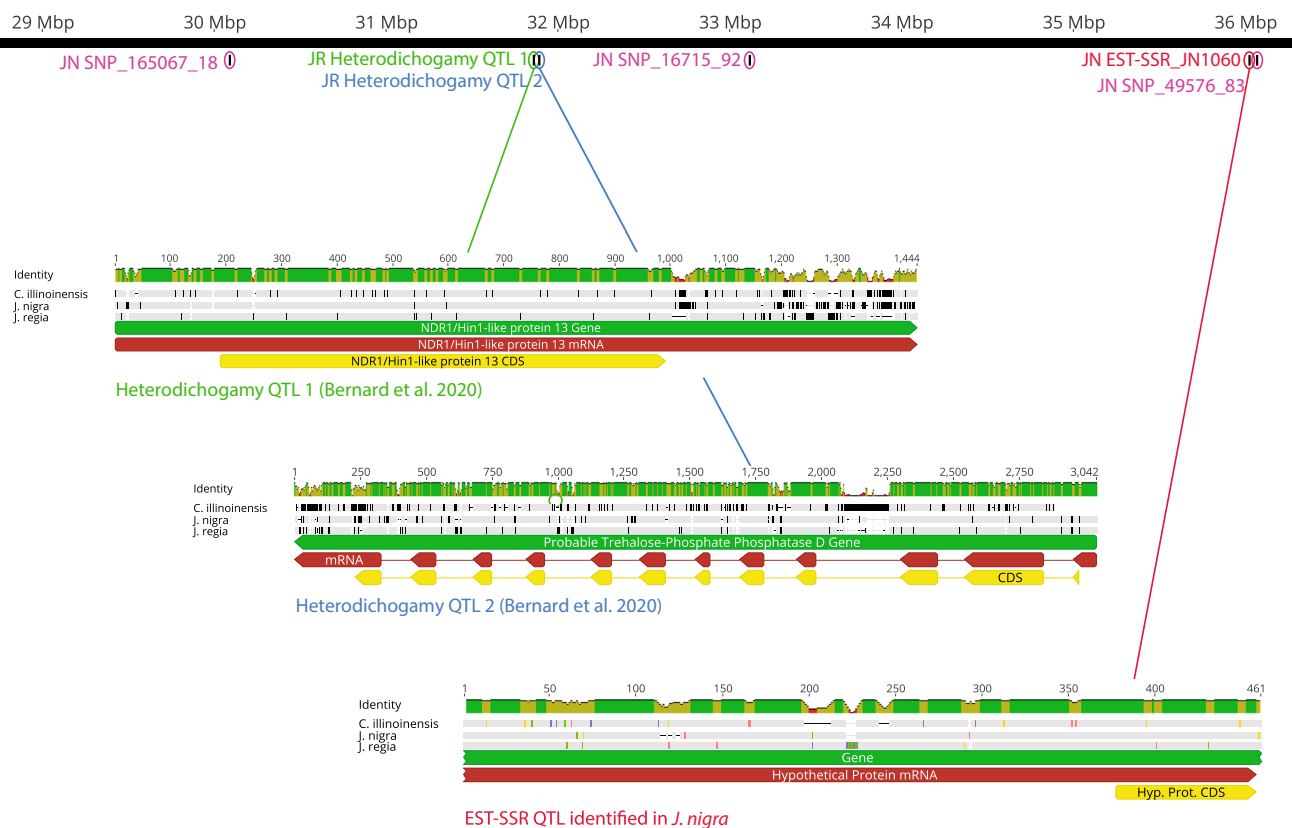
<sup>b</sup>Ninety-five percent confidence intervals were estimated using the 2-LOD drop method

<sup>c</sup>Percent explained variation

heterodichogamy as a covariate, one QTL was detected on LG11 for peak pistillate and staminate bloom. The QTL for pistillate bloom peaks at 13.5 cM at a LOD of 7.1, holding a 95% CI interval of 2.0–22.1 cM and explaining 18.4% of the

variation. Peak staminate bloom resided in a similar location, 14.5 cM (95% CI 0.0–19.1 cM), and a LOD score of 12. Interestingly, both flowering time QTL localized outside of the 95% CI for heterodichogamy. Additionally, when using

## *J. nigra* Markers Aligned to *J. regia* Chromosome 11



**Fig. 3** Local alignment of known *J. regia* heterodichogamy QTL with the detected *J. nigra* heterodichogamy QTL. This figure shows the proximity of the detected heterodichogamy QTL in *J. nigra* to two previously identified QTLs for heterodichogamy in *J. regia* (Bernard et al., 2020) on chromosome 11. Three local alignments of the

detected QTL regions between *J. regia* (Chr 11), *J. nigra* (Chr 11), and *Carya illinoensis* (Chr 07) showcase the broad transferability of EST-SSR markers across the Juglandaceae family. JR = *Juglans regia* JN = *Juglans nigra*

heterodichogamy as a covariate, the 95% CI of peak staminate bloom does not overlap with that of heterodichogamy, while the 95% CI of peak pistillate bloom overlaps with that of heterodichogamy marginally (21.1–22.1 cM). The QTL results for peak staminate bloom without using heterodichogamy as a phenotypic covariate detects what appears to be a major effect QTL on LG11, which illustrates the effect of not accounting for phenotypic covariance in this mapping population.

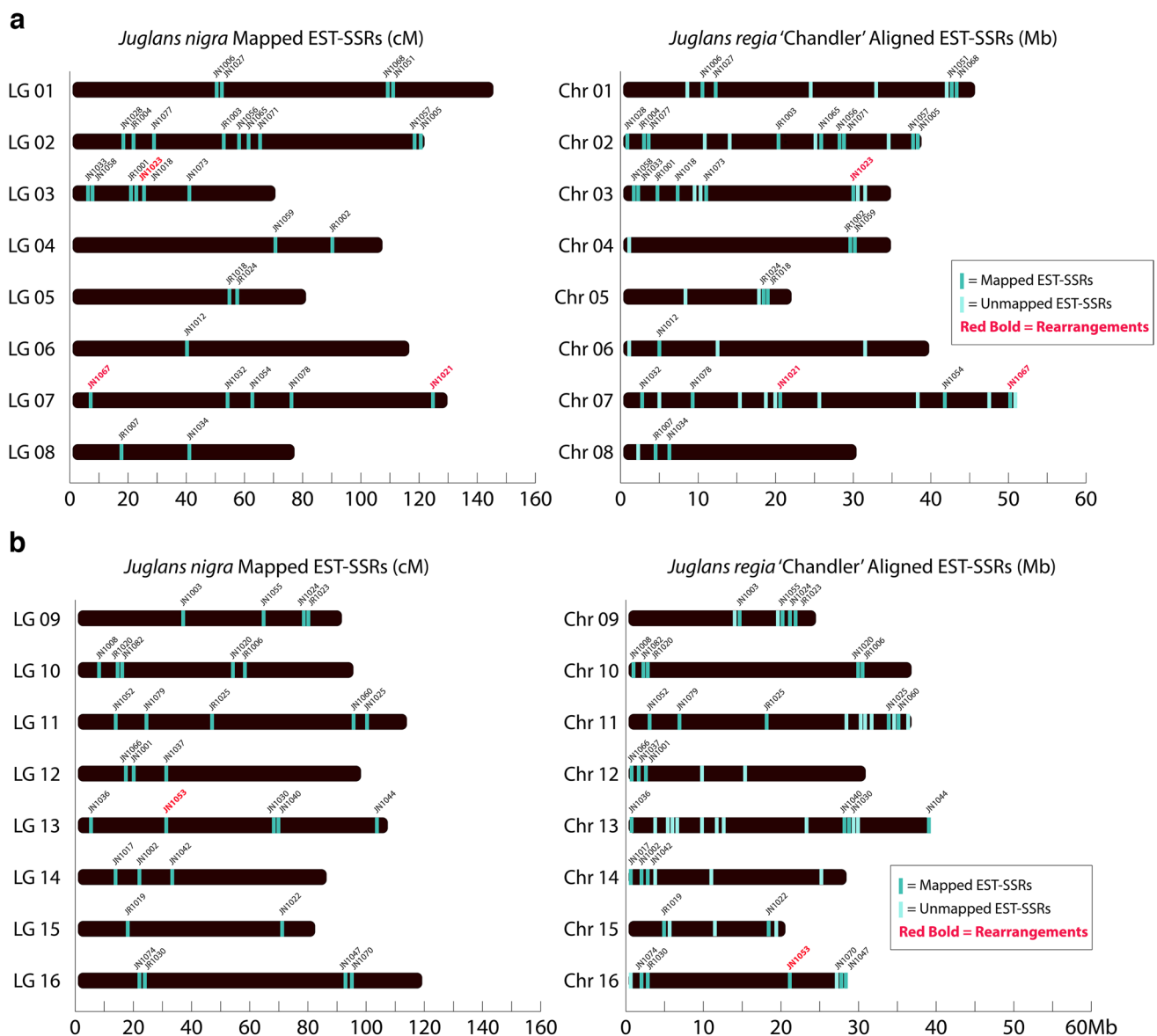
### Alignment with *J. regia*

Placement of the mapped EST-SSR markers by sequence alignment in *J. regia* and mapping in *J. nigra* suggests collinearity with some exceptions. All 62 of the EST-SSR markers placed on the linkage map were located in *J. regia* sequenced pseudomolecules, within the LGs and in the linear order expected, with some notable exceptions (Fig. 4, Online Resource 1). For example, ESTJN1053 which mapped to

LG13 was placed by sequence alignment on Chromosome 16 of *J. regia*. In the linkage map, ESTJR1023 on LG3 is flanked on one side by three EST-SSRs and by two EST-SSRs on the other, while it is located distant to the other five on Chromosome 3 in *J. regia*. The sequence positions in *J. regia* of the five EST-SSR markers mapped to *J. nigra* LG7 suggest a history of multiple rearrangements. On the other hand, eight EST-SSR markers mapped to LG2 are in the same order as their sequences on the *J. regia* Chromosome 2 pseudomolecule, with the ninth mapped marker on that LG only slightly misplaced. We also located the 56 EST-SSR marker sequences uninformative for genetic mapping in the “Sparrow” x “Schessler” population to *J. regia* pseudomolecules (Online Resource 1).

### EST-SSR gene annotations

Our EST-SSR marker gene annotations were gathered from the *J. regia* genome annotations using blastn on the NCBI



**Fig. 4** Alignment of select *J. nigra* linkage groups with *J. regia* chromosomes showcasing possible genomic rearrangements between species. The horizontal black bars on the left of the image represent the *J. nigra* linkage groups with units in centimorgans (cM). The horizontal black bars on the right represent identified EST-SSR markers aligned to the *J. regia* “Chandler” v2.0 genome with units

in megabases (Mb). Teal vertical bars represent mapped EST-SSR markers, while light blue vertical bars represent sequenced EST-SSR markers that did not map but aligned to the “Chandler” genome. Markers with red text indicate locations with potential genomic rearrangements. **A** LG/Chr 1–8, **B** LG/Chr 9–16

nucleotide database and covered a wide array of functions (Online Resource 2, Sheet 1). One hundred EST-SSRs of the 118 potentially mappable markers had e-values of zero. One mapped EST-SSR derived from *J. regia* (estJR1011) remains unannotated. Annotation using a predicted genome annotation for *J. nigra* (Trouern-Trend et al. 2019) was attempted by creating a custom blast database out of the annotated *J. nigra* scaffolds in Geneious and searched with

the blastn algorithm. Overall, the results using the predicted annotations were less complete and inconsistent with the results from the NCBI nucleotide database (despite a high prevalence of e-values of zero). Ninety-five EST-SSRs of the 118 potentially mappable markers had e-values of zero. Forty-three EST-SSRs remained unannotated (31 originating from *J. nigra*, and 12 from *J. regia*) (Online Resource 2, Sheet 2).

## Discussion

### Summary of findings

This investigation describes the first QTL mapping results from a 13-year-old mapping population of full siblings, a population that will be a fruitful source of gene discovery in *J. nigra*. We have located a major QTL for heterodichogamy, the two morph flowering system of the monocious Juglandaceae, and provide evidence for synteny with *J. regia*. We have also detected QTL for budbreak, and peak bloom dates for pistillate and staminate flowers. Our results add to the evidence for a major gene system controlling heterodichogamy, with protogyny dominant to protandry. We also show evidence for separate QTL influencing peak pistillate and staminate bloom linked to the QTL region for heterodichogamy.

We have shown that in *J. nigra*, intramorph protogynous mating occurs naturally, as shown previously in *J. ailantifolia*. We observed blooming date overlap in our two protogynous parents, as evident by multiyear observations and the production, by natural pollination, of the parentage-verified ‘Sparrow’ by ‘Schessler’ F<sub>1</sub> mapping population. Intramorph mating occurs because the interval between pistillate and staminate bloom varies considerably among protogynous individuals, with intervals among protandrous individuals varying less, at least in our mapping population. In assessments of heterodichogamy in *J. regia*, individuals with less than seven days of separation between pistillate and staminate bloom (−3 to +3) are considered homogamous rather than heterodichogamous (Bernard et al. 2020). The genetic control of heterodichogamy may involve several genes, one with a major dominant effect and others that influence days of separation. As our mapping population continues to mature and multiyear observations accumulate, we can begin to investigate the genetic mechanisms that contribute to days of separation within morphs. The location of different QTL regions for peak pistillate bloom, peak staminate bloom and heterodichogamy on LG11 in *J. nigra* is intriguing, but the functional meaning, if any, requires additional investigations.

The association and linkage mapping investigation on phenological traits in *J. regia* detected a highly associated SNP and a colocalized major QTL for budbreak on chromosome 1, associated SNPs on chromosomes 2, 4, 8, and 15, with minor QTL on chromosomes 6, 11, 12, and 14 (Bernard et al. 2020). In contrast, our QTL analysis detected highly significant but moderate to minor QTL on LG2 and LG11. A QTL investigation in pecan (*Carya illinoensis*) found evidence for synteny between the four walnut LD blocs reported by Marrano et al. (2019a, b), the highly associated *J. regia* budbreak QTL on chromosome

1 reported by Bernard et al. (2020), and locus *Bb-03* on chromosome 3 in pecan. Although our segregating population was somewhat smaller than those reported in *J. regia* and *C. illinoensis*, QTL of the effect sizes reported by both groups should have been detected in our investigation if present. We surmise that our mapping population may not be polymorphic for the QTL region involved. The male parent “Schessler” breaks bud an average of 15 days earlier than the female parent “Sparrow.” If both parents are homozygous but have different alleles for the linkage block involved, then all progeny will have the same heterozygous allelic configurations. The possibility that a given mapping population does not segregate for a trait of interest illustrates the value of capturing as much polymorphism as possible in multiple mapping populations and large-scale genome-wide association studies.

### Linkage map comparison to *J. regia*

The final consensus map contains 62 EST-SSR markers and 256 SNP markers with a total length of 1646 cM (Table 3, Online Resource 1). The average marker density was 4.2 cM, with 19 regions having gaps greater than 10 cM. The average gap length was 12.9 cM. For comparison, the *J. regia* “Chandler” x “Idaho” map used 312 F<sub>1</sub> progeny to construct individual SNP marker maps for each parent (Marrano et al. 2019b). The female map contained 1860 SNP markers with a length of 1034 cM. The male map contains 2607 SNP markers with a length of 1724 cM. The female and male maps had five regions with gaps greater than 10 cM. The linear placement of markers and the adjusted estimate of the cM distance between markers are statistical inferences dependent on the size of the mapping population, the type of structured population used, the extent of actual recombination in the mapping population, the number of informative markers used to detect recombinants, and the technical accuracy of the markers used to detect recombination. Thus, the comparison of different maps’ quality is meaningful only when anchored by a high-quality genome sequence or physical map, such as the EST-SSRs reported here. The similar locations of major phenological QTLs on LG11 suggests that our linkage map will be sufficient for drawing initial comparisons between *J. regia* and *J. nigra* marker-trait association studies.

### Heterodichogamy in *J. nigra*

In the UMCA repository, of those cultivars that are characterized, there are 46 protogynous and 14 protandrous individuals. Interestingly, the composition of protogynous and protandrous *J. regia* cultivars in California is the opposite, where most are protandrous due to the prevalent use of certain founder parents. Dichogamous habit does not provide

a direct benefit to improve orchard performance. However, protandry is correlated with early foliation, robust branch, and architecture in timber forms of *J. nigra* (Pang et al. 2021). Additionally, the character also represents useful variation that can be introduced into pedigrees to shift pistillate/staminate bloom time. The trait has not yet been used in this way for *J. nigra* breeding, but this utility is conceivable for pollinizer development as new black walnut cultivars are released, especially pollinizers for protogynous cultivars with early pistillate bloom dates. In pecan, heterodichogamy is an essential character for facilitating optimal orchard pollination (Bentley et al. 2019). Figure 1 displays budbreak and bloom times for nine protogynous and six protandrous cultivars, representing the range of phenological times. Based on recommendations for *J. regia*, pollen shed must occur several days in advance of female receptivity for ideal fertilization to occur. MAS would be very applicable to pedigrees focused on pollinizer development, assuming the crosses are between protogynous and protandrous individuals.

### QTL mapping of heterodichogamy and peak flower bloom

Heterodichogamy served as an ideal QTL target for linkage map validation. Segregation patterns in *Juglans spp.* and *C. illinoensis* suggested that a single, dominant locus confers the protogynous habit (Thompson and Romberg 1985; Gleeson 1982). The habit is prevalent across the *Juglandaceae* and other families in the *Fagales* (Endress 2020), and QTLs have been recently discovered in *J. regia* (Bernad et al. 2020) and *C. illinoensis* (Bentley et al. 2020). Heterodichogamy mapped to a 185 kb region on chromosome 4 in *C. illinoensis* and to a 19 kb region on chromosome 11 in *J. regia* (Bernardo et al. 2020; Bentley et al. 2020). In *J. regia*, a gene encoding a probable trehalose-phosphate phosphatase D was discovered in the 19 kb region. This enzyme is a member of a gene family that is highly expressed in *Jatropha*, which expresses delayed staminate flowering (Zhao et al. 2019). It is hypothesized that the enzyme is involved in sugar metabolic pathways involved in inducing flowering transition.

Opportunity for comparative analysis across closely related species is expanding with greater sequencing depth, genome-wide studies, and improved chromosome-scale reference genomes. While these technologies are increasingly more accessible, they are not immediately deployable in minor crops, crop relatives, or underutilized species due to funding limitations or available genetic resources. In the interim, linkage maps anchored with family- or genus-informative EST-SSR allow an opportunity for comparative mapping to assess co-localization of QTL with phenotype. In such cases, the genomic resources of major species can be leveraged to better understand trait molecular evolutionary biology. *Next steps.*

The continued development of high-quality genetic resources for *J. nigra* should be prioritized. Currently, the “Sparrow” draft genome, assembled from Illumina shotgun sequenced long-insert mate pair libraries, has been used to address basic evolutionary questions such as phylogenetics, examining for introgression, and studies of ecological niche differentiation (Stevens et al. 2018). However, long-read sequencing of “Sparrow” combined with improved genome scaffolding (i.e., Hi-C) would accelerate assembly to chromosomal-scale and greatly enhance the value of the genetic map reported here, allowing for more in-depth comparisons with other *Juglans* species that have high-quality genomes (Marrano, et al. 2020; Ning, et al. 2020).

Future QTL likely to be developed in the “Sparrow” x “Schessler” bi-parental mapping population can be identified by analyzing divergent phenotypes in its parents. Segregation in the progeny are likely for the following traits: nut size (small vs. large), anthracnose susceptibility (low vs. medium), season length to nut maturity (116–120 days vs. 130–142 days), alternate bearing (0.5–0.7 alternate bearing index vs. 0.7–0.8) shell thickness and possibly outer shell and inner shell thickness (based on photographic observations). Nut size is not an important trait for cultivar selection in *J. nigra* as it is not marketed to the consumer as an in-shell product. However, shell thickness is a critical phenotype influencing the ease of shelling and separation of kernel halves (Reid et al. 2004). Anthracnose resistance has also been identified as a priority trait for the UMCA *J. nigra* breeding program (Reid et al. 2004). Recent investigations of differential transcriptomic and proteomic responses of *J. regia* walnut fruit to anthracnose infection may provide candidate genes and proteins for investigation in *J. nigra* (Fang et al. 2021; Feng et al. 2021).

Modern genomic techniques for reduced representation sequencing and targeted genotyping should also be developed for *J. nigra*. They can be applied to increase the SNP genotype density in the “Sparrow” x “Schessler” mapping population, improve QTL accuracy, and used to perform GWAS in the UMCA black walnut germplasm repository (Bernard et al. 2021). As more molecular-trait associations are developed, KASP-like assays for targeted QTL validation in different genetic backgrounds will be possible (Bernard et al. 2020; Fleming et al. 2022; Kaur et al. 2020; Wang et al. 2018; Zhang et al. 2021).

## Materials and methods

### Mapping population parents

“Sparrow,” the female parent of the mapping population, is a wild origin selection from Lomax, IL (40.35° N latitude), first named in 1935. “Sparrow” is one of the most productive nut cultivars in the University of Missouri Eastern Black

Walnut Germplasm Repository, with a yield efficiency of 10.86 nuts per cm<sup>2</sup> trunk area and a mean alternate bearing index of 0.5 to 0.7 for the period of 2002–2009. “Sparrow” exhibits strong tolerance to anthracnose defoliation, exhibiting only 6.7% defoliation rate of late season foliage over three years. Anthracnose, caused by *Gnomonia leptostyla* ([Fr.] Ces. & De Not.), can lead to significant premature defoliation in susceptible black walnut trees. Severe defoliation has been observed to lead to poor nut filling and alternate bearing. “Sparrow” nuts are relatively small (18.0 g/nut), averaging 29.94% edible kernel, produced over a 116–120-day season.

“Schessler” is also a wild origin tree selected in Genoa, OH (40.20° N latitude) in 1937. In contrast to “Sparrow,” the yield efficiency for “Schessler” is less, with only 8.03 nuts produced per cm<sup>2</sup> trunk area and a less desirable mean alternate bearing index of 0.7 to 0.8 for the period of 2002–2009. Late-season defoliation due to walnut anthracnose in “Schessler” averaged 16.8% over 3 years, higher than the value (6.8%) typically seen in “Sparrow.” “Schessler” nuts are larger (22.5 g/nut) than “Sparrow,” averaging 26.07% edible kernel, produced over a 130–142-day season. “Schessler” trees are among the earliest trees to leaf out in the MU repository.

Observations of the parent trees over 5 years (2002–2006) indicate that both are protogynous, “Schessler” breaks bud an average of 15 days before “Sparrow,” and the pistillate flowers become receptive an average of 13 days before those of “Sparrow” (Fig. 1). While the Julian dates for budbreak and flowering dates may vary by several weeks across years due to environmental influences, in our orchards, “Schessler” always breaks bud and flowers before “Sparrow.” Both parents are protogynous and the flowering of pistillate and staminate flowers overlap (Fig. 1).

## Mapping population

The *J. nigra* mapping population consists of 301 F<sub>1</sub> progeny of the two cultivars, “Sparrow,” the seed parent, and “Schessler,” the pollen parent. These full-sib progenies were identified using a parentage analysis of 3009 seedlings grown from nuts borne in 2008 on a single “Sparrow” tree located in a clonal orchard consisting of 19 unique cultivars. Microsatellite marker analysis using 11 gSSRs developed from *J. nigra* for all of the trees in the orchard as well as the 3009 open pollinated progeny identified “Schessler” as the pollen parent of 301 individuals with > 99% exclusion probability calculated using Cervus v3.0 (Kalinowski et al. 2007; Woeste et al. 2002). The full-sib seedlings were potted, grown another year, then established in 2010 on 6-m centers in a planting adjacent to the repository at the Horticulture and Agroforestry Research Center in New Franklin, MO (39° 00′ 50″ N lat, 92° 46′ 12″ W long).

## Phenotyping and data analysis

Evaluation of the following four traits occurred in 2020 and 2012. Budbreak was additionally evaluated in 2016 and 2017. In each year, the orchard three times a week during April and May to record Julian date of budbreak, peak pistillate bloom, and peak staminate bloom. These traits were evaluated based on the descriptors for walnuts of the International Plant Genetic Resources Institute (1994). Budbreak is defined as when over 50% of terminal buds are enlarged, with split scales, and green leaves are inside. Peak pistillate bloom is an estimation of when over 50% of pistillate flowers have become receptive, and peak staminate bloom is an estimation of when over 50% of staminate flowers have reached pollen shed. Heterodichogamy is calculated by subtracting peak staminate bloom from peak pistillate bloom, to numerically represent the degree flowering overlap. This value was assigned to one of the five following categories of an ordinal scale: 1, very protandrous (< − 10 days); 3, Protandrous (− 10 to − 4 days); 5, homogamous (− 3 to + 3 days); 7, protogynous (+ 4 to + 10 days); 9, very protogynous (> + 10 days). Best linear unbiased predictions (BLUPs) were calculated for each phenological trait to estimate genotypic effect, removing environmental effects, using a mixed linear model with year as a fixed effect and genotype as a random effect:

$$P_{ik} = \mu + Y_i + g_k + e_{ik}$$

where  $P_{ik}$  refers to the observed phenotype of the  $k$ th accession in the  $i$ th year;  $\mu$  is the mean value of the trait;  $Y_i$  is the fixed effect of the  $i$ th year;  $g_k$  is the random effect of the  $k$ th genotype; and  $e_{ik}$  is the residuals of the model. BLUPs were calculated in the “R” package “lme4” (Bates et al. 2015).

We used an unpaired t-test to evaluate three phenotypic contrasts: the difference between the mean peak pistillate bloom dates of protogynous and protandrous individuals, the difference between the mean peak staminate bloom dates of protogynous and protandrous individuals, and the difference between the mean interval between staminate and pistillate bloom dates for protogynous and protandrous individuals.

## Sample collection

For ddRadseq, twig tissue was freshly collected in 2016 and stored in a − 80 °C freezer. Samples were ground in liquid nitrogen in a SPEX 6850 Freezer mill (Metuchen, NJ) and DNA extracted with a Norgen Biotech Plant/Fungi DNA extraction kits (Norgen Biotek Corp®, Thorold, ON). DNA was resolved by electrophoresis on agarose gels to check for band quality, and by a ThermoScientific NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, MA). For EST-SSR marker development, late spring leaves were

collected in 2017 and processed in the same manner as for twigs.

### ddRADseq sequencing and data pipeline

In 2016, 277 full-sib progenies were selected at random for double-digest reduced representation sequencing (ddRAD-seq) at BGI International (Cambridge, MA) using an Illumina HiSeq 2000 and 101 bp paired-end reads. Funding limitations prevented genotyping the full population. As no sequenced *Juglans* genome was available at the time to permit a tailored choice of restriction enzymes, we used *EcoRI* and *MseI*. The protocols and informatics pipelines were as previously described for a northern red oak (*Quercus rubra* L.) mapping population (Konar et al. 2017).

### EST-SSR development and data pipeline

Six hundred eighty *J. nigra* EST-SSRs were selected from the Hardwood Genomics database (<https://www.hardwoodgenomics.org/>). One hundred thirty-four markers were removed because they contained more than one SSR in the reported sequence. Fifty-seven markers were further removed because they did not match at least 80 bp of any *Juglans* in the NCBI database using the BLAST search algorithm. This resulted in 489 EST-SSR markers from *J. nigra*.

EST-SSR markers were also mined from a transcriptome for *J. regia* from the Hardwood Genomics database using MISA, the MicroSatellite identification tool (<http://pgrc.ipk-gatersleben.de/misa/>). A total of 1847 EST-SSRs were detected. One hundred thirty-eight markers contained more than one SSR in the transcript. Two hundred sixty-two were further removed because they did not match at least 80 bp of any *Juglans* samples in the NCBI database using the BLAST search algorithm. This resulted in 1451 EST-SSR markers from *J. regia*.

A total of 1940 EST-SSR sequences were provided to MYcroarray®, now Daicel Arbor Biosciences (Ann Arbor, USA)™, for bait design. Candidate sequences were then filtered to minimize cross-hybridization and size selected in silico. The 683 candidate sequences remaining were used to construct custom baits.

### Target capture and sequencing

The 160 progeny with the least amount of missing ddRAD-seq data, along with the parents “Sparrow” and “Schessler,” were selected for target capture. Sample libraries were prepared at the Bioinformatics and Genomics Core Facility at the University of Notre Dame. Samples were prepared using the NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina with the NEBNext® Multiplex Oligo Set (Dual-index) adapters (New England BioLabs® Inc., Ipswich,

MA). DNA fragmentation was performed enzymatically to account for an average bait length of ~ 177 bp. Libraries were then validated with the Qubit™ dsDNA H.S. Assay Kit (ThermoFisher Scientific, Waltham, MA) and Bioanalyzer DNA 7500 kit (Agilent Technologies, Santa Clara, CA). Parent libraries were pooled for target capture and evaporated to a final volume of 15.0 µL. F<sub>1</sub> libraries were pooled into random groups of 5 random samples and evaporated to a final volume of 15.0 µL. The bait-capture procedure was performed by the Genomics Core Facility at the University of Notre Dame, following the myBaits® experimental guide (Daicel Arbor Biosciences, Ann Arbor, MI).

### EST-SSR pipeline

The N.D. Genomics & Bioinformatics Core Facility constructed a custom pipeline for the EST-SSR sequence data. First this pipeline utilized, TRIMMOMATIC (Bolger et al. 2014) was to remove sequencing adapters, and the read quality was assessed with FASTQC. BWA (Li and Durbin 2009) was used to align the trimmed reads to the previously designed bait sequences. A custom suite of tools was developed to extract the actual base pair lengths of each EST-SSR and the sequence of the embedded SSR. These data were placed in a spreadsheet, and each allele was called manually. An allele was considered plausible if it had the highest read count (minimum of 10 reads) for that individual marker. A second allele was called if it had a read count (minimum of 10 reads) that was at least 20% of the read count of the first called allele. Markers were discarded if they had more than 15% missing data. Finally, all genotypes were converted to their appropriate JoinMap codes, and impossible genotypes (such as those receiving two alleles from the mother tree) were re-coded at missing data. EST-SSR markers with more than 54% missing data were removed, leaving 118 EST-SSR markers available for mapping.

### Linkage map construction and QTL analysis

Mapping was done in JoinMap 4.1 (Van Ooijen and Jansen 2013). All markers were evaluated with a chi-square goodness-of-fit test for the expected Mendelian distribution of the parental genotypes prior to mapping. The goodness-of-fit test is good practice because the maximum likelihood model for pairwise recombination estimation depends on the expected Mendelian segregation ratios for both markers. The cutoff value for the test statistic was 15 for ddRADtag markers and 60 for EST-SSR markers. The cutoff value was determined by the lowest test statistic value in which one of the expected classes was not recovered, given the marker type.

JoinMap constructs a maximum likelihood expression for estimating recombination for all pairwise recombination

estimates. Grouping is then done using the LOD value of the test for independence using the default parameter (LOD = 1.0), which resulted in 15 groups. We used a Monte Carlo maximum likelihood (ML) approach to generate an initial estimate of linkage within a group. As this approach must fit all of the markers in the group, poor nearest neighbor fit (how well a locus fits between its neighboring loci) and nearest neighbor stress (the difference between the observed recombination of the joint segment and its prediction of recombination independence) will generate impossibly long LGs. After examination and removal of the poorest fit and highest stress-causing markers, two-round regression maps for 14 of the groups were made using Kosambi correction with default parameters. Our ML maps revealed that one of the 15 groups had a disproportionately large map (5800 centimorgans (cM) vs. an average of 287 cM for the other groups). Further, this map exhibited two distinct subgroups of about 400 cM in length, separated by 3 SNP markers spanning thousands of cM. We used Geneious Prime v2020.2.3 (<https://www.geneious.com>) to utilize NCBI's blastn algorithm to compare the mapped black walnut EST-SSR sequences within the two 400 cM subgroups to the *J. regia* V2.0 genome. This strategy located two EST-SSR markers in one subgroup to *J. regia* chromosome 15 and other four EST-SSR markers in the other subgroup to *J. regia* chromosome 16. Based on these results, we excluded the intervening three SNP markers, leaving two subgroups which we mapped separately using two round regression mapping.

QTL mapping was done using the approximate multiple-QTL model (MQM) implemented in MapQTL® (Van Ooijen and Jansen 2013), with the map generated on the same population phenotyped for dates of budbreak, peak pistillate bloom, and peak staminate bloom. Significance thresholds were determined by permutation test (10,000 permutations). We used peak pistillate bloom date as a covariate for the budbreak QTL analysis to account for the phenotypic covariance between these traits in *J. nigra*. We used heterodichogamy as a covariate for the peak pistillate bloom date and the peak staminate bloom date for the same reason. Phenotypic covariance affects every individual in the analysis, a confounding factor that can be accounted for by using one phenotype as covariate for the phenotype under examination.

### Alignment with *J. regia*

We utilized Geneious Prime to use Bowtie2 on Medium Sensitivity to align the full suite of sequenced *J. nigra* EST-SSRs (mapped and unmapped) and mapped *J. regia* EST-SSRs to the *J. regia* “Chandler” v2.0 genome (Langmead and Salzberg 2012). The starting bp positions of the aligned markers in the *J. regia* genome were noted and put into a map notation format

readable in MapChart v2.32. This created maps that, instead of showing LGs in cM, shows the linear order of our EST-SSR markers in megabases (Mb) in the *J. regia* genome (Online Resource 1). The maps were visualized using MapChart v2.32 (Voorrips 2002).

### EST-SSR gene annotations

We chose the EST-SSR markers based on informativeness and PCR performance in both *J. nigra* and *J. regia*, and not on inferred gene function. However, it is reasonable to assume that highly conserved expressed sequences will have very similar functions within the same genus and even in the same family. Our EST-SSR marker gene annotations were gathered from the *J. regia* genome annotations using blastn on the NCBI nucleotide database and covered a wide array of functions (Online Resource 2). Annotation using a predicted genome annotation for *J. nigra* (Trouern-Trend et al. 2019) was attempted by creating a custom blast database out of the annotated *J. nigra* scaffolds in Geneious and our 118 potentially mappable EST-SSRs were used as the query with the blastn algorithm.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s11295-022-01580-9>.

**Acknowledgements** Special thanks are given to the University of Missouri Center for Agroforestry. Without their long-term curation of this population, this work would not have been possible. The authors would also like to thank Robert Stanley and Angelyn Hilton for reviewing the manuscript.

**Funding** This work was supported by the National Science Foundation (IOS1025974) grant to J. Romero-Severson and M. Coggeshall (Co-P.I.s) and J. Carlson (P.I.), and by the Hardwood Tree Improvement and Regeneration Center (HTIRC) at Purdue University. This work is currently supported by the University of Missouri Center for Agroforestry and the USDA–ARS Dale Bumpers Small Farm Research Center, Agreement numbers 58–6020–6–001 and 58–6020–0–007 from the USDA Agricultural Research Service.

**Data Availability** Data is archived in NCBI or will be made available upon request.

### Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Conflict of interest** The authors declare that AD is employed by Daicel Arbor Biosciences which manufactures the myBaits enrichment kit used in this study.

**Data archiving statement** Our raw sequence data for both our RRS data and EST-SSR bait capture data is being submitted to NCBI's Short Read Archive (SRA). Accession numbers will be provided

once available. Our genetic maps will be submitted to the TreeGenes database and an accession number will be provided on acceptance.

The custom code used to process the ddRADseq reads is on GitHub (<https://github.com/jsarro13/SSRTools>) and will be publicly released upon acceptance of the manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aradhya MK, Velasco D, Wang JR, Ramasamy R, You FM, Leslie C, Dandekar A, Luo MC, Dvorak J (2019) A fine-scale genetic linkage map reveals genomic regions associated with economic traits in walnut (*Juglans regia*). *Plant Breeding* 138(5):635–646
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48
- Bentley N, Grauke LJ, Klein P (2019) Genotyping by sequencing (GBS) and SNP marker analysis of diverse accessions of pecan (*Carya illinoensis*). *Tree Genet Genomes* 15(1):8
- Bentley N, Grauke LJ, Ruhlman E, Klein RR, Kubenka K, Wang X, Klein P (2020) Linkage mapping and QTL analysis of pecan (*Carya illinoensis*) full-siblings using genotyping-by-sequencing. *Tree Genet Genomes* 16(6):1–20
- Bernard A, Lheureux F, Dirlwanger E (2017) Walnut: past and future of genetic improvement. *Tree Genet Genomes* 14(1):1
- Bernard A, Barreneche T, Lheureux F, Dirlwanger E (2018) Analysis of genetic diversity and structure in a worldwide walnut (*Juglans regia* L.) germplasm using SSR markers. *PLOS ONE* 13(11):e0208021
- Bernard A, Marrano A, Donkpega A, Brown PJ, Leslie CA, Neale DB, Lheureux F, Dirlwanger E (2020) Association and linkage mapping to unravel genetic architecture of phenological traits and lateral bearing in Persian walnut (*Juglans regia* L.). *BMC Genomics* 21(1):203
- Bradbury D, Smithson A, Krauss SL (2013) Signatures of diversifying selection at EST-SSR loci and association with climate in natural Eucalyptus populations. *Mol Ecol* 22(20):5112–5129
- Coggeshall MV (2011) Black walnut: a nut crop for the Midwestern United States. *HortScience* 46:340–342
- Cruden RW (1988) Temporal dioecism: systematic breadth, associated traits, and temporal patterns. *Bot Gaz* 149(1):1–15
- Daniels DA, Nix KA, Wadl PA, Vito LM, Wiggins GJ, Windham MT, Ownley BH, Lambdin PL, Grant JF, Merten P, Klingeman WE (2016) Thousand cankers disease complex: a forest health issue that threatens Juglans species across the US. *Forests* 7(11):260
- Endress PK (2020) Structural and temporal modes of heterodichogamy and similar patterns across angiosperms. *Bot J Linn Soc* 193(1):5–18
- Fang H, Liu X, Dong Y, Feng S, Zhou R, Wang C, Ma X, Liu J, Yang KQ (2021) Transcriptome and proteome analysis of walnut (*Juglans regia* L.) fruit in response to infection by *Colletotrichum gloeosporioides*. *BMC Plant Biology* 21(1):1–16
- Feng S, Fang H, Liu X, Dong Y, Wang Q, Yang KQ (2021) Genome-wide identification and characterization of long non-coding RNAs conferring resistance to *Colletotrichum gloeosporioides* in walnut (*Juglans regia*). *BMC genomics* 22(1):1–17
- Fleming MB, Miller T, Fu W, Li Z, Gasic K, Saski C (2022) PpeXapF: High throughput KASP assays to identify fruit response to *Xanthomonas arboricola* pv. *pruni* (Xap) in peach. *PloS one* 17(2):p.e0264543
- Fukuhara T, Tokumaru S-I (2014) Inflorescence dimorphism, heterodichogamy and thrips pollination in *Platycarya strobilacea* (Juglandaceae). *Ann Bot* 113(3):467–476
- Gleeson SK (1982) Heterodichogamy in walnuts: inheritance and stable ratios. *Evolution* 892–902
- International Plant Genetic Resources Institute (1994) Descriptors for walnut (*Juglans* spp.). Bioversity International
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099–1106
- Kaur A, Ferguson L, Maness N, Carroll B, Reid W, Zhang L (2020) Spring freeze damage of pecan bloom: a review. *Horticulturae* 6(4):82
- Kefayati S, Ikhsan AS, Sutyemez M, Paizila A, Topçu H, Bükücü ŞB, Kafkas S (2019) First simple sequence repeat-based genetic linkage map reveals a major QTL for leafing time in walnut (*Juglans regia* L.). *Tree Genet & Genomes* 15(1):13
- Kimura M, Seiwa K, Suyama Y, Ueno N (2003) Flowering System of Heterodichogamous *Jugland aillanthifolia*. *Plant Species Biol* 18(2–3):75–84
- Kimura MK, Goto S, Suyama Y, Matsui M, Woeste K, Seiwa K (2012) Morph-specific mating patterns in a low-density population of a heterodichogamous tree *Juglans Aillanthifolia*. *Plant Ecol* 213(9):1477–1487
- Konar A, Choudhury O, Bullis R, Fiedler L, Kruser JM, Stephens MT, Gailing O, Schlarbaum S, Coggeshall MV, Staton ME, Carlson JE (2017) High-quality genetic mapping with ddRAD-seq in the non-model tree *Quercus rubra*. *BMC Genomics* 18(1):417
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357
- Leonarduzzi C, Spanu I, Labriola M, González-Martínez SC, Piotti A, Vendramin GG (2016) Development and Characterization of Three Highly Informative EST-SSR Multiplexes for *Pinus halepensis* mill. and their Transferability to Other Mediterranean Pines. *Plant Mole Biol Report* 34(5):993–1002
- Lloyd DG, Webb CJ (1986) The avoidance of interference between the presentation of pollen and stigmas in angiosperms I. Dichogamy New Zealand J Botany 24(1):135–162
- MacDaniels LH (1933) Nut growing in New York State, vol. Bulletin 573. Cornell University Agricultural Experiment Station, Ithaca, pp 24
- Marrano A, Martínez-García PJ, Bianco L, Sideli GM, Di Pierro EA, Leslie CA, Stevens KA, Crepeau MW, Troggio M, Langley CH, Neale DB (2019a) A new genomic tool for walnut (*Juglans regia* L.): development and validation of the high-density Axiom™. *J regia* 700K SNP genotyping array. *Plant Biotechnol J* 17(6):1027–1036
- Marrano A, Sideli GM, Leslie CA, Cheng H, Neale DB (2019b) Deciphering of the genetic control of phenology, yield, and pellicle color in Persian walnut (*Juglans regia* L.). *Frontiers in Plant Science* 10(1140)
- Marrano A, Britton M, Zaini PA, Zimin AV, Workman RE, Puiu D, Bianco L, Pierro EAD, Allen BJ, Chakraborty S, Troggio M (2020) High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. *GigaScience* 9(5):giaa050
- Martínez-García PJ, Crepeau MW, Puiu D, Gonzalez-Ibeas D, Whalen J, Stevens KA, Paul R, Butterfield TS, Britton MT, Reagan RL, Chakraborty S et al (2016) The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding

- for the biosynthesis of non-structural polyphenols. *Plant J* 87(5):507–532
- McCleary T, McAllister M, Coggeshall M, Romero-Severson J et al (2013) EST-SSR markers reveal synonymies, homonymies and relationships inconsistent with putative pedigrees in chestnut cultivars. *Genet Resour Crop Evol* 60(4):1209–1222
- McGranahan G, Leslie C (2012) Walnut. In *Fruit breeding*. Springer, pp 827–846
- McKenna JR, Coggeshall MV (2018) The genetic improvement of black walnut for timber production. In *Plant Breeding Reviews*. John Wiley & Sons, pp 263–289
- Meier N, Coggeshall MV, Webber JB, Bunting E, Revord RS (2022) Genetic variation among 54 eastern black walnut cultivars for phenological and morphological traits. *J Am Soc Hortic Sci* 147(5):281–289
- Michler CH, Woeste KE, Pijut PM (2007) Black walnut. In: KC (ed) *Genome mapping and molecular breeding in plants*. Springer, Heidelberg, pp 189–198
- NASS (2020) Noncitrus Fruits and Nuts 2019 Summary. USDA, National Agricultural Statistics Service
- Ning D-L, Wu T, Xiao LJ, Ma T, Fang WL, Dong RQ, Cao CL (2020) Chromosomal-level assembly of *Juglans sigillata* genome using Nanopore, BioNano, and Hi-C analysis. *GigaScience* 9(2):giaa006
- Pang K, Woeste KE, Saunders MR, McKenna JR, Mickelbart MV, Jacobs DF, Michler CH (2021) Rapid growth in clonal *Juglans nigra* L. is most closely associated with early foliation, robust branch architecture, and protandry. *For Ecol Manag* 499:119590
- Pollegioni P, Olimpieri I, Woeste KE, De Simoni G, Gras M, Malvolti ME (2013) Barriers to interspecific hybridization between *Juglans nigra* L. and *J. regia* L. species. *Tree Genetics & Genomes* 9(1):291–305
- Pollegioni P, Woeste KE, Chiocchini F, Del Lungo S, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2015) Ancient humans influenced the current spatial genetic structure of common walnut populations in Asia. *PLoS ONE* 10(8):e0135980–e0135980
- Pollegioni P, Woeste K, Chiocchini F, Del Lungo S, Ciolfi M, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti ME (2017) Rethinking the history of common walnut (*Juglans regia* L.) in Europe: Its origins and human interactions. *PLoS one* 12(3):e0172541
- Reid W, Coggeshall MV, Hunt KL (2004) Cultivar evaluation and development for black walnut orchards. United States Department Of Agriculture Forest Service General Technical Report Nc 243:18
- Renner SS (2001) How common is heterodichogamy? *Trends Ecol Evol* 16(11):595–597
- Roor W, Konrad H, Mamadjanov D, Geburek T (2017) Population differentiation in common walnut (*Juglans regia* L.) across major parts of its native range—insights from molecular and morphometric data. *J Hered* 108(4):391–404
- Santos C, Nelson CD, Zhebentyayeva T, Machado H, Gomes-Laranjo J, Costa RL (2017) First interspecific genetic linkage map for *Castanea sativa* x *Castanea crenata* revealed QTLs for resistance to *Phytophthora cinnamomi*. *PLoS ONE* 12(9):e0184381
- Sideli G, McAtee P, Marrano A, Allen BJ, Brown PJ, Butterfield TS, Dandekar AM, Leslie CA, Neale DB (2020a) Genetic analysis of walnut (*Juglans regia* L.) pellicle pigment variation through a novel, high-throughput phenotyping platform. *G3 Genes Genomes Genetics* 10(12):4411–4424
- Sideli GM, Marrano A, Montanari S, Leslie CA, Allen BJ, Cheng H, Brown PJ, Neale DB (2020b) Quantitative phenotyping of shell suture strength in walnut (*Juglans regia* L.) enhances precision for detection of QTL and genome-wide association mapping. *PLOS ONE* 15(4):e0231144
- Stevens KA, Woeste K, Chakraborty S, Crepeau MW, Leslie CA, Martínez-García PJ, Pui D, Romero-Severson J, Coggeshall M, Dandekar AM, Kluepfel D (2018) Genomic variation among and within six *Juglans* species. *G3: Genes Genomes Genet* 8(7):2153–2165
- Talbert TJ (1942) Nut tree culture in Missouri, vol. bulletin 454. University of Missouri, College of Agriculture, Agricultural Experiment Station, pp 32
- Thompson TE, Romberg LD (1985) Inheritance of heterodichogamy in pecan. *J Hered* 76(6):456–458
- Torokeldiev N, Nelson CD, Zhebentyayeva T, Machado H, Gomes-Laranjo J, Costa RL (2018) Genetic diversity and structure of natural *Juglans regia* L. populations in the southern Kyrgyz Republic revealed by nuclear SSR and EST-SSR markers. *Tree Genet Genomes* 15(1):5
- Trouern-Trend A, Falk T, Zaman S, Caballero M, Neale DB, Langley CH, Dandekar A, Stevens KA, Wegrzyn JL (2019) Comparative genomics of six *Juglans* species reveals patterns of disease-associated gene family contractions. *Plant J* 102(2):410–423
- Vahdati Kourosh, Arab MM, Sarikhani S, Sadat-Hosseini M, Leslie CA, Brown PJ (2019) Advances in Persian walnut (*Juglans regia* L.) breeding strategies. In: Al-Khayri JM, Jain SM, Johnson DV (eds) *Advances in plant breeding strategies: Nut and beverage crops*, Volume 4. Springer International Publishing, Cham, pp 401–472
- Van Ooijen JW, Jansen J (2013) *Genetic mapping in experimental populations*. Cambridge University Press
- Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93
- Wang R, Liu Y, Isham K, Zhao W, Wheeler J, Klassen N, Hu Y, Bonman JM, Chen J (2018) QTL identification and KASP marker development for productive tiller and fertile spikelet numbers in two high-yielding hard white spring wheat cultivars. *Mol Breeding* 38(11):1–12
- Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. *J Hered* 93(1):58–60
- Zeven AC, Zhukovsky PM (1975) *Dictionary of cultivated plants and their centres of diversity*, Centre for Agricultural Publishing and Documentation. Wageningen Natura Montenegro 3:2004
- Zhang Y, Li T, Geng Y, Wang Y, Liu Y, Li H, Hao C, Wang H, Shang X, Zhang X (2021) Identification and development of a KASP functional marker of TaTAP46-5A associated with kernel weight in wheat (*Triticum aestivum*). *Plant Breed* 140(4):585–594
- Zhao ML, Ni J, Chen MS, Xu ZF (2019) Ectopic expression of *Jatropha curcas* TREHALOSE-6-PHOSPHATEPHOSPHATASE J causes late-flowering and Heterostylous phenotypes in Arabidopsis but not in *Jatropha*. *Int J Mol Sci* 20:2165
- Zhu Y, Yin Y, Yang K, Li J, Sang Y, Huang L, Fan S (2015) Construction of a high-density genetic map using specific length amplified fragment markers and identification of a quantitative trait locus for anthracnose resistance in walnut (*Juglans regia* L.). *BMC Genomics* 16(1):614