Inferring the Genetic Basis of Sex Determination from the Genome of a Dioecious Nightshade

Meng Wu,[1] David C. Haak (iD),[2] Gregory J. Anderson,[3] Matthew W. Hahn,[1,4] Leonie C. Moyle,[1] and Rafael F. Guerrero (iD)*,[4,5]

[1]Department of Biology, Indiana University, Bloomington, IN, USA
[2]School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA
[3]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
[4]Department of Computer Science, Indiana University, Bloomington, IN, USA
[5]Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA

***Corresponding author:** E-mail: rfguerre@ncsu.edu.
Associate editor: Michael Purugganan

Abstract

Dissecting the genetic mechanisms underlying dioecy (i.e., separate female and male individuals) is critical for understanding the evolution of this pervasive reproductive strategy. Nonetheless, the genetic basis of sex determination remains unclear in many cases, especially in systems where dioecy has arisen recently. Within the economically important plant genus *Solanum* (~2,000 species), dioecy is thought to have evolved independently at least 4 times across roughly 20 species. Here, we generate the first genome sequence of a dioecious *Solanum* and use it to ascertain the genetic basis of sex determination in this species. We de novo assembled and annotated the genome of *Solanum appendiculatum* (assembly size: ~750 Mb scaffold N50: 0.92 Mb; ~35,000 genes), identified sex-specific sequences and their locations in the genome, and inferred that males in this species are the heterogametic sex. We also analyzed gene expression patterns in floral tissues of males and females, finding approximately 100 genes that are differentially expressed between the sexes. These analyses, together with observed patterns of gene-family evolution specific to *S. appendiculatum*, consistently implicate a suite of genes from the regulatory network controlling pectin degradation and modification in the expression of sex. Furthermore, the genome of a species with a relatively young sex-determination system provides the foundational resources for future studies on the independent evolution of dioecy in this clade.

Key words: sex chromosome, sex evolution, genome assembly, dioecy, *Solanum*.

Introduction

Dioecy—the presence of separate female and male individuals in a species, usually accompanied by dramatic phenotypic differences between the sexes—is a substantial, evolutionarily potent source of intraspecific variation. In species with genetic sex determination, these differences between males and females extend to their genomes, which show genomic differentiation in and around the sex-determining region (or SDR) (Charlesworth 2002, 2013). Theory for the origin and emergence of these genomic signatures of dioecy—including sequence divergence in the SDR (Charlesworth and Charlesworth 1978; Rice 1987)—is well developed. According to the most prevalent model, genetically determined dioecy starts with mutations at two loci, one mutation resulting in male sterility and one mutation leading to female sterility; these two mutations are expected to be tightly linked within a single genomic region. The SDR is later formed by the suppression of recombination between these two loci. After establishment, the SDR can continue to expand (i.e., encroach on neighboring genomic regions), increasing the amount of the genome that is differentiated between the sexes (Otto et al. 2011). Although some evidence is consistent with this "two-locus" model for the evolution of dioecy (see below), this pathway may not be the only way to achieve genetic sex determination (Henry et al. 2018). Accordingly, it remains unclear whether the origin and early evolution of dioecy unfolds in predictable and generalizable ways across independent instances of the transition from hermaphroditism to dioecy.

Empirical data from a diversity of dioecious systems can provide critical information for understanding the underpinnings of this transition, and for identifying common genomic features of genetic sex determination. However, although dioecy is extremely common in animal species (where it is more often termed "gonochorism"), its origins are due to a few ancient events that are shared across major taxa, making analyses of the early emergence and evolution of dioecy challenging in animal systems. In contrast, although only 5–6% of

angiosperm species have separate male and female individuals (Renner 2014), dioecy is phylogenetically widespread, occurring across 43% of angiosperm families (Renner 2014). It has been estimated that dioecy arose at least 100 times independently within the flowering plants (Charlesworth 2002; Renner 2014). This repeated emergence of dioecy in plants makes them ideal for the study of genetic sex determination and the early stages of sex-chromosome evolution, in ways that ancient and static animal systems cannot match (Charlesworth 2015).

Indeed, recent genomic analyses in young dioecious plant systems have made substantial progress toward evaluating the nature of the emerging SDR and uncovering the genetic factors and molecular mechanisms that regulate sex determination. For instance, sex differences in frequency of k-mers (sequence motifs of varying lengths) have been used to identify sex-specific genomic sequences and to identify the SDR in persimmon, kiwifruit, and date palm tree (Akagi et al. 2014, 2018; Torres et al. 2018). Genomic approaches, combined with data from more classical analyses, have been used several times to test the two-locus model for the evolution of dioecy in plants (Charlesworth and Charlesworth 1978), with evidence for this model found in species such as *Silene*, papaya, asparagus, and kiwifruit (Wang et al. 2012; Kazama et al. 2016; Harkess et al. 2017; Akagi et al. 2019). In other species, however, a single gene seems to be sufficient for the expression of maleness and the repression of female development (as in persimmon) (Akagi et al. 2014, 2016).

Despite this emerging clarity, several challenges still arise in studying the genetic basis of dioecy in angiosperms. Because dioecy in many plants is not ancient, locating the SDR based on obvious signatures (such as large-scale genomic differentiation) used for many animal sex chromosomes (Charlesworth 2013, 2015) might not be appropriate (but see clear exceptions in *Silene*, papaya, among others). Compounding this problem, plant genomes are often rich in transposable elements, which can contribute to the fast accumulation of repetitive sequence around the nonrecombining SDR (Li et al. 2016); this can both obscure simple signatures of genomic differentiation expected to be associated with this region and can lead to difficulties in the de novo assembly and reconstruction of this region. Other challenges are biological. For instance, the genetic mechanisms involved in flower development are numerous and complex (involving various genes, small RNAs, epigenetic effects, and environment interactions) (Martin et al. 2009; Nag and Jack 2010; Adam et al. 2011; Song et al. 2013; Akagi et al. 2016; Bräutigam et al. 2017; Murase et al. 2017), making it difficult to pinpoint the specific loci responsible for sex determination in any particular case. In addition, although several studies have made important contributions to understanding the genetic basis underpinning the evolution of dioecy, most currently examined plant systems also have high levels of sexual dimorphism, amplifying the potential for multiple correlated phenotypic changes to confound or mask the causal loci involved in the initial emergence of sex differentiation.

Dioecious species in the genus *Solanum* present the opportunity to study the recent and repeated evolution of dioecy in this speciose and economically important plant clade. Although most of the nearly 2,000 species in the genus are hermaphroditic, there are a number of cases of dioecy, andromonoecy (in which male and hermaphrodite flowers occur separately on the same plant), and even sexually fluid species (Anderson 1979; Symon 1979, 1981; Anderson and Levine 1982; Levine and Anderson 1986; Whalen and Costich 1986; Anderson and Symon 1989; Anderson et al. 2015; McDonnell et al. 2019). Of these, there are 15–20 clearly dioecious *Solanum* species (Anderson et al. 2015), and their distribution across the phylogeny of the genus suggests that dioecy has independently evolved at least four times (i.e., dioecious species appear in four different sections: one in the potato group, one in Dulcamaroid, two in Geminata, and 14 in Leptostemonum; Knapp et al. 2004; Anderson et al. 2015). In all the dioecious species, female flowers still produce anthers and the majority of the anthers on the female flowers also include pollen (Anderson et al. 2015). The pollen grains produced by females, despite having high viability (i.e., a metabolically active cytoplasm inferred from stain assays) and many of the same internal constituents as the pollen from male flowers, do not function in the usual way. They lack the typical three apertures that are important for emergence of the pollen tube (that is, they are inaperturate; Levine and Anderson 1986) (fig. 1). Nonetheless, the pollen in these female flowers plays another key role in *Solanum*: since there is no floral nectar (Anderson and Symon 1985), the inaperturate pollen is the only reward for the bee pollinators. Moreover, this consistent transition to inaperaturate pollen in female flowers suggests that independent transitions to dioecy within *Solanum* might often involve convergent phenotypic changes, at least for the emergence of females. Furthermore, there are currently no genomic analyses of dioecious *Solanum* species, and nothing is known of the mechanistic genetic basis of dioecy in this genus.

Thus, our goal was to address the mechanisms underpinning the emergence of dioecy in one such *Solanum* species. *Solanum appendiculatum* is the sole dioecious species in the Anarrhichomenum section of *Solanum*, which is estimated to be only 4 My old (Echeverría-Londoño et al. 2020). There is no evidence of sex chromosome divergence in the karyotypes of *S. appendiculatum* (Bernardello and Anderson 1990), and it is unknown whether males or females are heterokaryotypic (i.e., whether this is a XX/XY or ZZ/ZW sex-determination system). Consistent with a very recent origin of dioecy ($<$4 My), this species shows very little sexual dimorphism: female flowers are morphologically hermaphroditic but functionally pistillate due to the nongerminable inaperturate pollen, whereas the male flowers never set fruit (fig. 1) (Anderson 1979). Previous studies of the early development of pollen in *S. appendiculatum* showed that pollen from the female flowers develops a primexine (a cellulose layer that initiates the pollen wall formation) at the physical position that would otherwise develop apertures (Zavada and Anderson 1997). Intine (consisting primarily of cellulose and pectin) then accumulates and thickens the pollen wall at these locations, resulting in the failure to develop germination pores (Zavada and Anderson 1997). Although the breeding system of this species
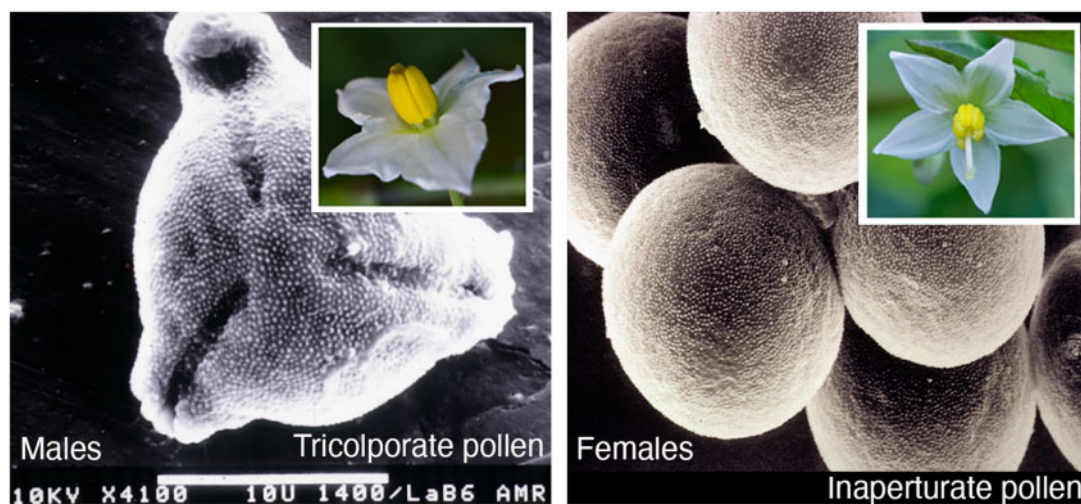
Fig. 1. Pollen produced in male and female flowers of *Solanum appendiculatum*. On the left, male pollen is tricolporate, typical of the genus. On the right, inaperturate pollen produced by females is shown (lacking pores). Flowers of each sex are shown in the upper inset.

(sensu Neal and Anderson 2005) is well understood, including the structural basis of functional male sterility in females, the genetic basis of sex determination is unknown.

In this study, we generated a high-coverage genome assembly of *S. appendiculatum* by adopting a hybrid assembly strategy using PacBio long reads and Illumina short reads. We paired this with RNA-seq data from male and female bud and flower tissue. Based on the genome annotation in this new assembly, we assessed three lines of data for genomic signatures of sex determination: 1) we investigated gene family dynamics through comparative analyses with seven other annotated (nondioecious) genomes in the Solanaceae; 2) we identified sex-biased gene expression in floral bud tissue and mature flowers; and, 3) we identified candidate sex-specific regions based on the location of sex-specific *k*-mers in the genome assembly. Interestingly, we uncovered pectin-related genes in all of these analyses. Pectin is associated with several sex-specific functions in plants (Micheli 2001), including a direct functional role for pectin production in pollen wall and germination pore development (Gou et al. 2012; Jiang et al. 2013). Therefore, we propose that the putative mechanism of the transition to inaperturate pollen in females of *S. appendiculatum* is a change in pectin regulation in female flowers. This genome assembly, and the putative SDR and candidate loci underpinning sex determination that emerge from it, also provide a clear framework for future studies of sex determination more broadly in *Solanum*.

Materials and Methods

Plant Sampling and Sequencing

Individuals from two accessions (population locations) of *S. appendiculatum* (denoted accession nos. 670 and 716 by G.J.A.), originally collected from natural habitats in Mexico, were grown and maintained in the greenhouse at Indiana University. Accessions were founded by seeds of a single open-pollinated plant, so individuals are expected to be ap-proximately half-sibs within accession. The original collection data and voucher information for these accessions are provided in Supplementary Material online. We collected young leaves from male and female individuals (three of each from accession 670, two of each from accession 716) and extracted DNA using the Qiagen DNeasy Plant Mini Kit. We treated samples with RNAase-A and used 100 μl of elution buffer. From each of these samples, we sequenced genomic DNA using the Illumina sequencing-by-synthesis technology (2×150 bp using HiSeq). Additionally, we obtained long-read sequences from one male and one female (670-1190 and 670-34) using the PacBio platform (sequencing carried out by Novogene Hong Kong).

To characterize gene expression profiles, we collected buds (1 day prior to anthesis) and mature flowers from 12 individuals (three males and three females from each accession) on the same day. Additionally, we collected young leaves from the reference female (670-1190) from which we extracted RNA for genome annotation purposes. Tissues were ground using the Tissue-lyser (Qiagen), RNA was extracted using the Qiagen RNeasy Plant mini-kit and brought to a final concentration of 70–200 ng/μl. A subset of RNA samples was checked for quality by agarose gel electrophoresis and visualized by ethidium bromide staining. Sample quality was further evaluated using the Agilent 2200 RNA TapeStation system before library creation. Stranded, paired-end libraries of total RNA were generated for each sample using Illumina Truseq Stranded mRNA sample preparation kits. Short-read DNA and RNA quality control, library preparation, and pooling were performed by the Indiana University Center for Genomics and Bioinformatics.

Flow Cytometry

Fresh early-leaf samples were collected for three female and three male plants and shipped overnight on ice to Virginia Tech for flow cytometry. Because the predicted genome size was expected to be near 1 Gb, we used *Capsicum chinense*

(~3.2 Gb; Kim et al. 2017) as an internal standard. Leaf samples were prepared following Galbraith et al. (1983): in brief, 0.5 g of fresh leaf material from each species was placed in a petri dish and nuclei were released by chopping with a razor blade under 0.5 ml ice-cold DeLaat's buffer (de Laat and Blaas 1984). After chopping, another 0.5 ml of ice-cold DeLaat's buffer (total volume 1 ml) was used to wash the containing cell constituents into a 30-μm filter to separate nuclei from chopping debris. To concentrate the nuclei, samples were centrifuged under 4 °C for 5 min at 492 g, and the total volume reduced by removing the top aqueous solution, leaving a consistent 200 μl of solution with the soft pellet. Nuclei staining was achieved by adding 1:1 v/v staining solution containing propidium iodide (50 μg/μl), RNase (50 μg/ml), and ß-mercaptoethanol (1.1 μl/ml). After the addition of staining solution samples were gently mixed by inversion and then incubated in the dark at room temperature for 20 min. After allowing the incorporation of propidium iodide, samples were then kept in the dark and moved to a cooler with ice (~4 °C) while being processed in the Flow Cytometry Resource Laboratory at Virginia Tech. Relative fluorescence was measured with the FL2 detector, and DNA content was quantified with FL2-area (integrated fluorescence) and comparison of relative peak means between the unknown and internal standard (Baldwin and Husband 2013).

Genome Assembly and Quality Assessment

We followed a hybrid assembly pipeline, combining short and long reads to assemble and annotate the genome (similar to our previous work; Wu et al. 2019). We implemented the MaSuRCA v3.2.2 (Zimin et al. 2017) pipeline to generate the genome assembly, in which Illumina paired-end reads from one male plant (121× coverage from 670-1190) and one female plant (116× coverage from 670-34) were used together to trim and correct low base-call-accuracy long-reads generated by PacBio sequencing from the same two plants (19× coverage from the male 670-1190 and 21× coverage from the female 670-1190). The MaSurCA assembler has been shown to work well with heterozygous genomes, splitting highly divergent regions into separate contigs while attempting to combine regions with up to approximately 6% divergence (Zimin et al. 2013).

Genome size was estimated before assembly was carried out, based on the k-mer abundance distribution separately from the male and female reads using GenomeScope (Vurture et al. 2017), with a k-mer length of 25 bp and max k-mer coverage of 10,000. After initial assembly, all assembled scaffold sequences were aligned against bacterial, archaea, fungal, and human databases to remove potential contaminants in our assembly, using the DeconSeq tool v0.4.3 (Schmieder and Edwards 2011).

We evaluated the completeness of the genome assembly using 1,515 plant near-universal single-copy orthologs within BUSCO v3 (Simão et al. 2015). To provide an estimate of assembly accuracy at the nucleotide level, we calculated a quality score for every position in the genome assembly using the program Referee (Thomas and Hahn 2019). Referee compares the log-ratio of the sum of genotype likelihoods for the genotypes that contain the reference base (e.g., [A, A], [A, T], [A, C], and [A, G] for reference base "A") versus the sum of those that do not contain the reference base (e.g., [T, T], [T, C], [T, G], [C, C], [C, G], and [G, G] for reference base "A"). The input used in the Referee calculation was obtained from the output pileup file from ANGSD (Korneliussen et al. 2014), which precalculated genotype likelihoods at each base of the genome assembly. Here, two genotype likelihood scores for every position in the genome assembly were calculated separately based on either the BAM file of aligned Illumina reads from the male (770-34) or the BAM file of aligned the Illumina reads from the female (670-1190).

Repeat and Gene Annotation

We followed the "Repeat Library Construction—Advanced" steps from the MAKER-P pipeline (Campbell et al. 2014) to generate a species-specific repeat library. Briefly, the LTR retrotransposon (LTR-RT) library was constructed using LTRharvest (Ellinghaus et al. 2008) and LTR-retriever (Ou and Jiang 2018). Miniature inverted transposable elements (MITEs) were detected by MITE-Hunter (Han and Wessler 2010). Other repetitive sequences were identified with RepeatModeler (http://www.repeatmasker.org/RepeatModeler/, last accessed March 30, 2021). All the detected repetitive elements above were combined into the S. appendiculatum-specific library. RepeatMasker (Smit et al. 2015) was then used to mask all repeat elements in the assembled genome by searching for homologous repeats in the species-specific library.

Gene models were predicted using three classes of evidence: RNA-seq data, protein homology, and ab initio gene prediction (see supplementary text, Supplementary Material online), which was performed within the MAKER-P pipeline (Campbell et al. 2014). All information on gene structures from the three different classes was synthesized to produce final gene annotations. We extracted relatively high-confidence genes according to evidence-based quality, requiring an annotation evidence distance (AED, which measures the goodness of fit of an annotation to the RNA/protein-alignment evidence supporting it) score <0.5. To assign functions to genes, we followed the pipeline AHRD (https://github.com/groupschoof/AHRD, last accessed November 28, 2020) to automatically select the most concise, informative, and precise functional annotation (see supplementary text, Supplementary Material online). For each gene, the associated gene ontology (GO) annotations were assigned according to the predicted protein domains (http://www.geneontology.org/external2go/interpro2go, last accessed November 28, 2020). To investigate the putative genomic locations of predicted genes in the genome, we performed whole-genome synteny alignment using Satsuma v3.1.0 (Grabherr et al. 2010), and recorded the genes unambiguously associated with a single identified syntenic region in the tomato (S. lycopersicum) genome (version ITAG3.20) (The Tomato Genome Consortium 2012).

Gene Family Analyses

To investigate changes in gene family size specifically in *S. appendiculatum*, we determined the expanded or contracted gene families along this branch in a phylogenetic tree. We obtained the protein-coding sequences of several Solanaceae species with high-quality genome annotations, including *Solanum lycopersicum*, *Solanum tuberosum*, *Jaltomata sinuosa*, *Nicotiana attenuata*, *Petunia axillaris*, and *Petunia inflata*. Gene families from these species were inferred by the program OrthoFinder (Emms and Kelly 2015), as were the phylogenetic relationships among species. Divergence times were retrieved from previous estimates within the Solanaceae (Särkinen et al. 2013). Rapidly evolving gene families along the branch leading to *S. appendiculatum* were determined using the program CAFE v3.1 (Han et al. 2013) with a *P* value cutoff of 0.01.

Gene Expression Analyses

For the RNA-seq data, trimmed reads from each library were mapped against the assembled genome using HISAT v2.1.0 (Kim et al. 2015), and the resulting SAM files were then converted to sorted BAM files using SAMtools v0.1.19 (Li et al. 2009). Reads assigned to exonic regions of genes were counted, and read counts of each annotated gene were calculated using FeatureCounts (Liao et al. 2014). The raw read count table was used as input for the program edgeR (Robinson et al. 2010). For downstream analysis, we removed genes with low expression by requiring each gene to have >1 read count per million (CPM) mapped reads in at least two samples (of any accession, sex, or tissue type). Trimmed mean of *m*-value (TMM) normalization was performed to eliminate composition biases between different libraries (Robinson and Oshlack 2010). Based on gene expression patterns, we also performed multidimensional scaling analysis in edgeR (Robinson et al. 2010) using the function "PlotMDS." The distances among samples were estimated using the average of the absolute value of the log2-fold expression changes among genes.

Differential expression analysis was performed separately for flower buds and mature flowers using edgeR (Robinson et al. 2010). We fit an additive model (expression $\sim$ accession $+$ sex) to control for the underlying differences among accessions, using the *glmfit* function. The genes that consistently displayed differential expression for a given pairwise sex comparison were identified using the function *glmLRT*. Genes were only considered to be significantly differentially expressed between the two sexes at a false discovery rate (FDR) <0.05 (Benjamini and Hochberg 1995) and fold change (FC) >2.

Search for Putative Sex-Determination Region

We used a *k*-mer approach to find putative sex-determination region sequences (Akagi et al. 2014, 2018). We searched for sex-specific *k*-mers—defined as 30-bp sequences detected in all samples of one sex but completely absent in the samples of the other—in the DNA short reads of six males and females (the same data set used in variant calling). We counted *k*-mer frequency in each sample using

Jellyfish v2.2.9 (Marçais and Kingsford 2011). To estimate the number of sex-specific *k*-mers expected by chance, we also counted private *k*-mers in 100 random combinations of six individuals from our data set. For each of these pseudosamples, we drew individuals within accession (three from each), and recorded *k*-mers present in the six chosen individuals and absent from the remaining six.

We complemented this approach by mapping the sex-specific *k*-mers to the assembled genome, guiding our inference in two ways. First, we prioritized genomic regions (10 kb windows) with high counts of sex-specific *k*-mers; we expect spurious *k*-mers to be roughly uniformly distributed across the genome, so this approach should find sex-specific regions. Second, candidate genomic regions can show patterns that are informative about the evolution and age of the sex-determination system. If the sex-determination region is highly diverged (e.g., if there is an old nonrecombining region), most sex-specific *k*-mers will map to contig(s) with large coverage differences between the sexes. For instance, in an evolutionarily old XX/XY system, male-specific *k*-mers map to Y-contigs where nearly no X-chromosome reads map, resulting in close to null coverage from females in that region. By contrast, there will be no difference in read-depth between the sexes if the sex-determining polymorphism is not in a nonrecombining region, or if such a region evolved recently.

We further characterized candidate genomic regions with standard population genetic statistics. To investigate sequence divergence between two sexes, we calculated measures of sequence differentiation (d_{XY} and F_{ST}) along with genetic diversity (π) in 10-kb windows (using the python scripts from https://github.com/simonhmartin/genomics_general, last accessed March 30, 2021). We filtered out low-quality sites (genotype quality <30 or read depth <5 in each individual) and used only windows with at least 1,000 sites after filtering. We calculated divergence between sexes within each accession and averaged the two values to obtain a single estimate per window.

All scripts written for this manuscript, as well as the MaSuRCA parameter files for the assembly, are available at http://doi.org/10.5281/zenodo.4495713 (last accessed March 30, 2021).

Results

Following the workflow (supplementary fig. S1, Supplementary Material online), we generated the first genome assembly of a dioecious species in *Solanum* (table 1). The genome of *S. appendiculatum* is estimated to be 671.8 Mb based on the *k*-mer frequency distribution, which is consistent with the estimated size from flow cytometry (750 $\pm$ 14 Mb for females and 742 $\pm$ 6 Mb for males; $n = 3$ for each sex). The size of the genome assembly is approximately 751.9 Mb, which includes 3,643 scaffolds with an N50 length of approximately 920.8 kb. In the assembly, approximately 96.6% of 1,515 BUSCO plant universal single-copy orthologous genes were found to be complete. 99.6% of sites in the assembly were well supported by either male or female Illumina reads with a quality score higher than 20

Table 1. Summary of the *Solanum appendiculatum* Genome Assembly.

Assembly features	
Genome size estimated (Mb)	671.83
Assembly size (Mb)	751.93
Number of scaffolds	3,643
Scaffold N50 length (kb)	920.78
GC contents	35.78%
Plant_CEGs (BUSCO)[a]	92.6 + (4.0 + 1.0)
Bases with quality score >20	99.57%
Repeat annotation	
Total (Mb)	497.56 (66.17%)
Gypsy (Mb)	254.87 (33.90%)
Copia (Mb)	26.98 (3.59%)
Unknown LTR-RTs (Mb)	90.79 (12.06%)
Gene annotation	
Number of protein-coding genes	35,731
Genes supported by RNA-seq	82.08%
Mean CDS length (bp)	1,265.56
Number of exons per gene	6.17

[a]Plant_CEGs (Clusters of Essential Genes) shows the percentage of complete single-copy orthologs plus the percentage of duplicated orthologs and fragmented orthologs.

(supplementary table S1, Supplementary Material online), suggestive of low base-calling error in our assembly. Base calls in the assembly were slightly better supported by female reads than male reads (99.4% vs. 97.7%) at a quality threshold of 20 as determined by separate runs of the Referee program (Thomas and Hahn 2019) (supplementary table S1, Supplementary Material online).

The genome of *S. appendiculatum* is highly repetitive, consisting of 497.6 Mb of repetitive sequence (66.2% of the assembly; supplementary table S2, Supplementary Material online). Consistent with other Solanaceae genomes, *Gypsy* LTR-RTs are the most abundant repeat elements in the *S. appendiculatum* genome (254.9 Mb; comprising 33.9% of the assembly). To annotate protein-coding genes, we used the MAKER-P pipeline, finding 35,731 high-confidence genes with AED scores <0.5, which is similar to the gene number (35,768) annotated in the domesticated tomato genome (The Tomato Genome Consortium 2012). The gene structure of annotated genes (e.g., CDS lengths, exon numbers, and GC contents) was also similar to the genes annotated in other Solanaceae genomes (table 1). Among high-confidence genes, we found 82.1% of their exonic regions are supported by at least one RNA-seq read, and 64.3% of them have expression levels with TPM (transcript per million) >1 in at least one sample of our investigated flower or leaf tissues. Together, the large scaffold size, the high coverage of the plant conserved single-copy genes, the low base-calling error, and the well-supported gene annotation all indicate a high-quality assembly for this species.

Pectin-Related Gene Families Have Evolved Dynamically in *S. appendiculatum*

We identified 228 rapidly expanded gene families and 75 rapidly contracted gene families on the branch leading to *S. appendiculatum* (supplementary tables S5 and S6, Supplementary Material online). As expected from previous studies in plants, many of those rapidly evolving gene families are involved in stress-related response, such as NBS-LRR disease-resistance proteins. Of the 228 rapidly expanding gene families in *S. appendiculatum*, we found five gene families that are functionally related to pectin degradation and modification. These included genes in the rhamnogalacturonate lyase, pectin lyase, invertase/pectin methylesterase inhibitor, and pectin-acetylesterase families (supplementary table S7, Supplementary Material online). In this last expanded family, there are two female-biased genes (*sapp52355* and *sapp52361*), as identified in our RNA-seq data (above). Lineage-specific expansion of pectin-related gene families was also found in other investigated Solanaceae genomes, but a smaller number of genes were duplicated in those species relative to *S. appendiculatum* (supplementary table S7, Supplementary Material online). In addition to pectin-related genes, we also detected two rapidly expanded gene families (OG0001468 [three genes gained] and OG0001481 [five genes gained]) and one rapidly contracted gene family (OG0000179 [three lost]) annotated as "plant self-incompatibility protein S1 family."

Sex-Biased Gene Expression Is Greatest in Mature Flowers

We tested for differential gene expression among sexes, separately in flower buds and in mature flowers. In buds, only 16 genes were identified to be significantly sex-biased (fig. 2A and supplementary table S3, Supplementary Material online), whereas 95 significantly sex-biased genes were detected in mature flowers (fig. 2B and supplementary table S4, Supplementary Material online). Almost two-thirds (58) of the differentially expressed genes in mature flowers are female-biased, with higher expression in the female flowers (fig. 2B and supplementary table S4, Supplementary Material online). This female bias is unlikely to be caused by differential read mapping onto a female-biased genome assembly, since we found no association between male–female sequence divergence and differences in gene expression (supplementary fig. S2, Supplementary Material online).

Among the genes showing sex-biased expression in mature flowers, we found several that may be functionally associated with phenotypic differences between males and females. In particular, we identified eight female-biased genes that are functionally related to pectin metabolism, which is involved in the production of germination pores on pollen (see Discussion). These eight genes include genes encoding a pectin acetylesterase family protein, a pectin lyase-like protein, and an invertase/pectin methylesterase inhibitor. The frequency of pectin-related genes is significantly higher in the set of sex-biased genes (Fisher's Exact $P = 3.17 \times 10^{-5}$) and female-biased genes (Fisher's Exact $P = 1.05 \times 10^{-6}$), compared with pectin-related genes as a fraction of all genes (416 out of 35,731 annotated genes). Several of the female-biased genes also appear to be pistil-specific (i.e., found only in female reproductive tract or ovary: the "pistil"). In particular, the gene *SlINO* has female-biased expression, and is a known
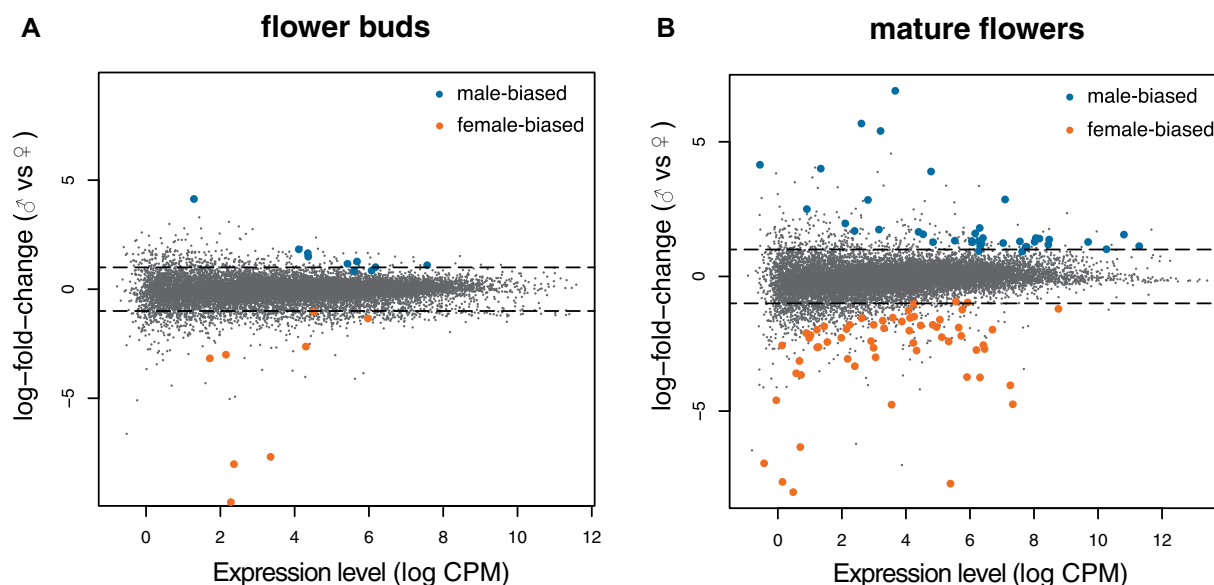
Fig. 2. Sex-biased gene expression pattern at two different flower development stages of *Solanum appendiculatum*. Differentiation gene expression between male and female flowers in (*A*) flower buds, and (*B*) mature flowers.

pistil-specific *YABBY* transcription factor in domesticated tomato (Ezura et al. 2017).

Candidate Sex-Determination Regions Suggest an XX/XY System

Based on a search for sex-specific sequences in six male and six female individuals, we identified approximately 11,000 female-specific and approximately 6,000 male-specific *k*-mers. Although the absolute *k*-mer count is larger for females, we found no evident excess of female-specific *k*-mers relative to total library size (supplementary table S8 and fig. S3, Supplementary Material online). Relative *k*-mer counts in both sexes are roughly within the expectation from randomized sets (supplementary fig. S3, Supplementary Material online). We found a small percentage of *k*-mers that could not be mapped to the genome assembly (3.74% for males, and 3.96% for females), but the similar fractions of unmapped *k*-mers for each sex suggest that there is no substantial sex bias in the assembly and that the sex-determination region is likely to be part of it (fig. 3A).

The distribution of sex-specific *k*-mers across the genome showed the first meaningful genomic difference between males and females (fig. 3B). We found a clear accumulation of male-specific content in a handful of 10-kb windows, whereas female-specific *k*-mers were uniformly spread throughout the genome (following a pattern similar to that of our randomized pseudosamples). In fact, 37% of male-specific *k*-mers were mapped to the top five windows, compared with only 18% in females. Moreover, the top five windows in males are clustered in two scaffolds (figs. 3 and 4), whereas only two of the top female windows are contiguous (supplementary fig. S4, Supplementary Material online). Two scaffolds harboring putative sex-determining polymorphisms (scf14997 and scf15476) were also recovered by two complementary searches for sex-specific regions (see Supplementary Material online). In particular, assembled contigs from PacBio

reads carrying male-specific *k*-mers were mapped to these same regions (supplementary tables S10 and S11, Supplementary Material online).

In these male-specific windows, we found other patterns typical of an XX/XY sex-determination system. Heterozygosity was uniformly higher in males, consistent with greater sequence divergence between the X and Y alleles present in males. We found increased relative and absolute divergence between males and females, but this observed divergence is not large enough to result in differences in read coverage (fig. 4). The read depth, divergence, and heterozygosity patterns in the female-specific regions were less clear, suggesting that the *k*-mer accumulation found is spurious (supplementary fig. S4, Supplementary Material online). Together, these results suggest that *S. appendiculatum* has an XX/XY sex-determination system.

We annotated 11 genes in the top male-specific regions (including 30 kb around the candidate windows; supplementary table S9, Supplementary Material online). Some of these annotated genes have been previously implicated in flower development and sex determination. Notably, we found three invertase/pectin methylesterase inhibitors (PMEIs; genes *sapp25116*, *sapp25117*, and *sapp25118* on scf14997), which could play a role in regulating pectin degradation or modification in pollen grains. The three PMEIs have sequence identity >90% and are located next to each other within a single syntenic block, consistent with recent tandem duplication events. In the window with the largest accumulation of male-specific *k*-mers (scf15476), we also found a transcription factor belonging to the R2R3 MYB superfamily; genes in this family commonly play regulatory roles in a wide array of processes in plants (including tissue development).

Discussion

We generated the first genome for a dioecious species within the genus *Solanum*, to assess the early emergence and genomic signatures of sex-differentiation and sex determination.
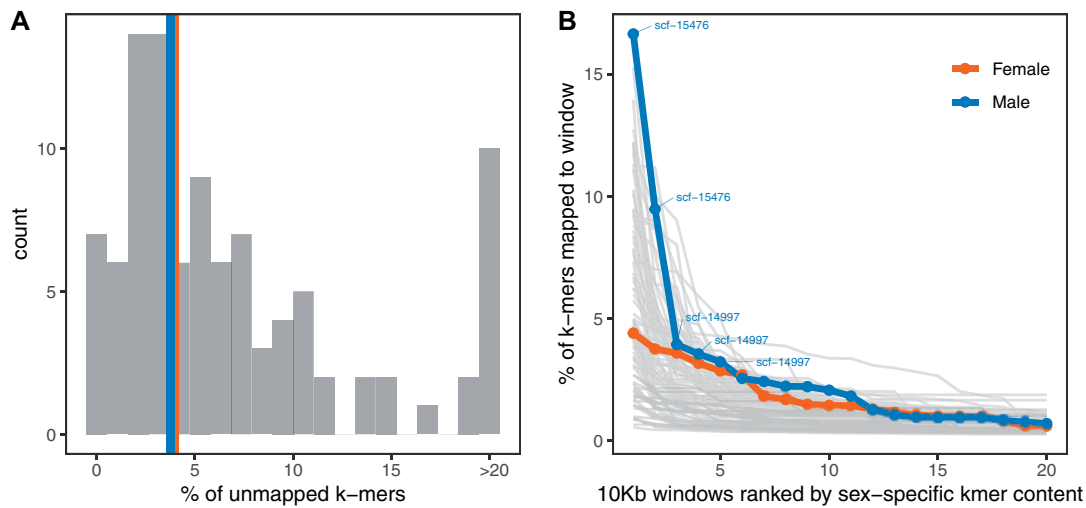
Fig. 3. (A) Distribution of private *k*-mers not found in the genome (unmapped) for 100 random pseudosamples of six out of 12 individuals (gray bars). Vertical lines show the observed value for female-specific (orange) and male-specific *k*-mers (blue). (B) The percent of sex-specific *k*-mers found in the twenty 10-kb windows with highest sex-specific *k*-mer content. Male-specific *k*-mers (blue) are accumulated largely in two scaffolds (top five annotated), whereas the female-specific (orange) more closely resembles the 100 random samples (gray).
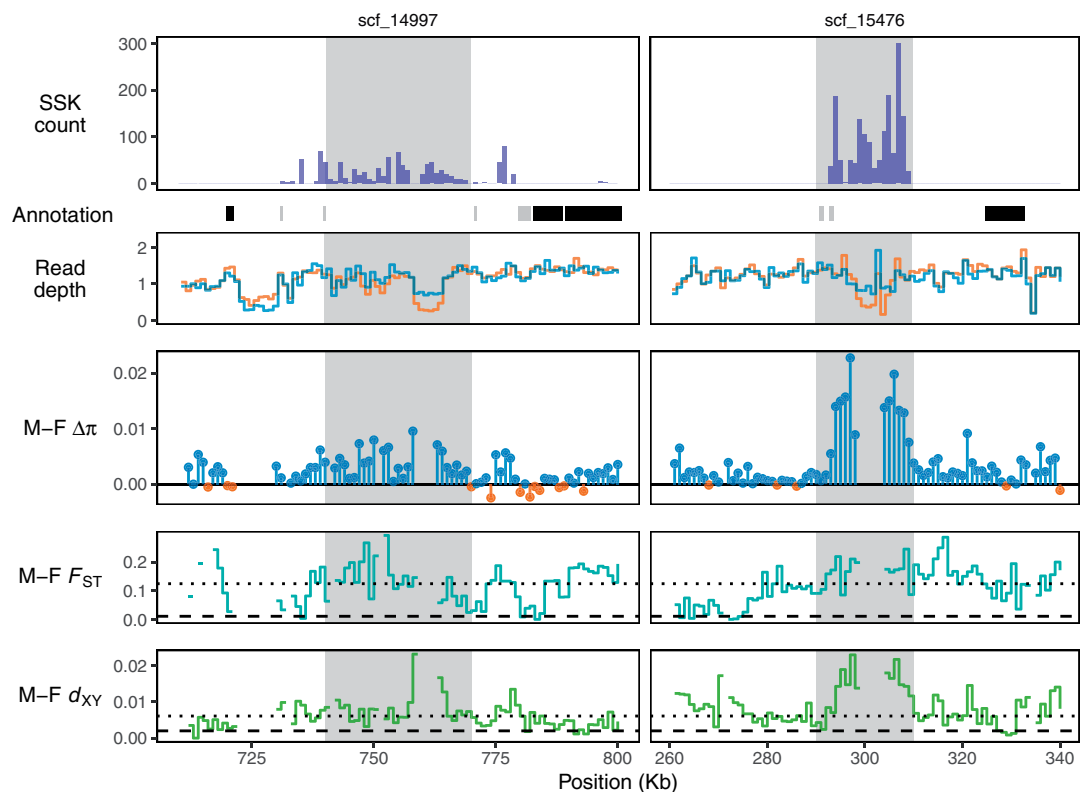
Fig. 4. Two genomic scaffolds that contain windows enriched with male-specific sequences (shaded light gray areas). Top row: Sex-specific *k*-mer count histogram. Below, rectangles are annotated genes in each scaffold (black: expressed in floral tissue but with no differential expression between sexes; gray: no evidence of expression in floral tissue; none of these genes showed significant sex-biased expression). Second row: Normalized read depth in male (blue) and female samples (orange). Third row: Difference in heterozygosity level (π) between males and females. Positive values indicate higher sequence diversity in males. Fourth row: Relative divergence (F_{ST}) between males and females. Last row: Absolute sequence divergence (d_{XY}) between males and females.

To do so, we assembled a high-quality genome, took a *k*-mer approach to find sex-linked genomic regions, and carried out an RNA-seq experiment of floral tissues to find genes involved in sex determination and sexual dimorphism. We found that dioecious *S. appendiculatum* appears to have a recently evolved sex-determination region and that males are likely

to be the heterogametic sex. Indeed, the patterns of male–female sequence divergence we observed do not indicate the presence of a large nonrecombining region containing genes involved in sex determination. Moreover, the specific loci associated with sex-differentiation suggest that the evolution of dioecy in this system involved changes in the regulation of pectin synthesis and degradation, including in specific phenotypic transitions observed in functionally female flowers. This genome, and the associated candidate genes, represents a valuable genomic resource for the continued investigation of recent transitions to dioecy within *Solanum*.

Limited Sex-Biased Gene Expression and Few Sex-Associated Regions Are Consistent with Recent Evolution of Sexual Dimorphism

We found a very modest amount of sex-biased gene expression in flower buds, and larger but still delimited sex differences in the expression profiles of mature flowers. Given that sex-specificity of gene expression is expected to accumulate with time since the origin of sexual dimorphism (Ellegren and Parsch 2007), the observation that few genes show sex-biased expression is consistent with a young sex-determination system. This very modest genomic and transcriptomic divergence between the sexes is consistent with the subtle morphological differentiation between male and female flowers, which is among the least pronounced in the dioecious nightshades (Anderson et al. 2015).

For mature flowers, sex-biased genes more commonly had higher expression in females than in males (fig. 2B). This finding contrasts with another species with a recently evolved sex-determining region–the garden asparagus (Harkess et al. 2015)—likely because of developmental differences in sex expression between the two systems. In asparagus, anther development is arrested before microspore meiosis in female flowers (Caporali et al. 1994), thus genes associated with later pollen development are expected to be expressed only in males (Harkess et al. 2015). In contrast, in *S. appendiculatum* female flowers develop mature pollen, but fail to deposit primexine at the apertural regions (Zavada and Anderson 1997). Our observation of more female-biased genes in *S. appendiculatum* is therefore consistent with this maintenance of both functional styles (female reproductive parts) and active production of (inaperturate) pollen (Levine and Anderson 1986) in female flowers, and seems to indicate some loss of function of female reproductive parts in male plants. This possible loss of function, however, is not reflected in the morphology of male flowers, which have complete female reproductive parts (albeit with much shorter styles; Anderson 1979; Anderson and Levine 1982).

Regulation of Pectin as a Potential Mechanism for the Formation of Aperturate Pollen

Identification of candidate genes playing potential feminizing or masculinizing effects is important to understand sex determination in this recently evolved dioecious species. Collectively, three different approaches in this study—gene family dynamics, sex-biased expression, and sex-specific *k*-

mers—detected a set of loci distinctive to *S. appendiculatum*. Some of these are likely unrelated to this species' transition to dioecy, and some others are possibly associated with general physiological consequences of this breeding system transition rather than directly involved in sex-differentiation and sex-determination per se. For instance, our gene family analysis detected a contraction of the self-incompatibility protein S1 family specifically in *S. appendiculatum*. Because the evolution of dioecy dramatically reduces the possibility of self-fertilization, this transition might be expected to relax selection to maintain functional self-incompatibility genes; similar losses of self-incompatibility proteins has also been observed in other Solanaceae species that have undergone breeding system transitions (e.g., to self-compatibility; Wu et al. 2019). Nonetheless, among the genetic changes detected, it is striking that all three of our different approaches detected pectin-related genes in association with sexual differentiation in *S. appendiculatum*, including pectin acetylesterases (PAE), pectin lyase-like proteins (PLL), and pectin methylesterase inhibitors (PMEI). Our finding is particularly intriguing as pectin synthesis and regulation is known to play important roles in pollen wall development, and in pollen function more broadly. Pectin consists of homogalacturonan (HG), which can be methyl- and acetyl-esterified (Wu et al. 2018), and pectin polysaccharides are critical components of the pollen wall. Mutants in genes encoding pectin polysaccharide synthetic and degrative enzymes—including pectin methylesterase (PME), polygalatcturonase (PG), PAE, and PLL—often show defective primexine, intine, or other pollen wall structures (Shi et al. 2015; Wu et al. 2018). Strikingly, in *Nicotiana* (Solanaceae), transgenic mutants of one pectin acetylesterase gene, *PAE1*, exhibit the loss of germination pores on the surface of the pollen grains (Gou et al. 2012)—a very similar phenotype to the inaperturate pollen observed in the female flowers of *S. appendiculatum*. The overexpression *PAE1* in transgenic tobacco results in severe male sterility by affecting the germination of pollen grains and the growth of pollen tubes (Gou et al. 2012).

Other pectin-associated proteins are also implicated in numerous functional roles in pollen tube germination and growth, including via coordinated regulation between PMEs and their inhibitors—PMEIs (Mollet et al. 2013). For instance, PME is important for the generation of methyl esterified HG in the apical zone of growing pollen tubes, which provides sufficient plasticity for sustaining growth (Cheung and Wu 2008). The removal of methyl ester groups by PME may allow the pectin-degrading enzymes, such as PLL or PG, to cleave the HG backbone, which may affect the rigidity of the cell wall (Gaffe et al. 1994; Micheli 2001). It has been proposed that the pollen cell might maintain a closely regulated level of PME activity, via regulation by PMEIs, in order to maintain the equilibrium between strength and plasticity in the apical cell wall (Bosch and Hepler 2005, 2006). For example, silencing of the *PME1* gene in tobacco (Bosch and Hepler 2006), and suppression of PMEI *At1g10770* in *Arabidopsis* (Zhang et al. 2010), both result in slowed pollen tube growth.

In addition to detecting sex-specific expression of PAE, we also found three PMEIs in a candidate sex-determining region (scf14997) in *S. appendiculatum*. The arrangement and relationship between these putative sex-determining genes are consistent with them being recent duplications, similar to what has been found in other dioecious plants (Harkess et al. 2017; Akagi et al. 2018). Although the specific function of these genes is not yet known, the general roles of PMEIs, PAE, and other related proteins in the formation and function of pollen suggests some possible models for the emergence of sex-specific pollen functions in the two sexes of *S. appendiculatum*. For example, it is possible that these PMEI copies influence the differential (sex-specific) expression patterns of downstream pectin-related genes in mature flowers, including PAE, thereby inhibiting or initiating the feminizing effect (i.e., inaperturate pollen) observed in female flowers. This process could also involve other tightly linked genes: the same syntenic block contains a gene coding for a *LOB* domain protein (*sapp25115*), the *Arabidopsis* ortholog of which (*AT1G06280*) is specifically expressed during tapetum and microspore development in the anthers (Oh et al. 2010; Zhu et al. 2010). Other differentially expressed genes also have clearly relevant functions. For example, the pyruvate dehydrogenase E1 component subunit alpha (*sapp29734*) was differentially expressed between males and females in the mature flower; pyruvate dehydrogenase catalyzes the early steps of sporopollenin biosynthesis, a major component of the exine layer of pollen grains (Jiang et al. 2013).

Although pectin-related genes are promising candidates for the expected male-sterilizing step in the evolution of dioecy, it is possible that they are downstream of a master regulator of sex determination. For instance, a MYB-like transcription factor similar to that found in scf15476 (gene *sapp39069*) has been implicated in the determination of sex in *Asparagus officinalis* (Murase et al. 2017), and the knockout of its putative ortholog causes male sterility in *Arabidopsis thaliana* (Zhu et al. 2008). Although the *sapp39069* transcription factor could be a regulator of sex, the R2R3 MYB superfamily has been shown to have an extreme diversity of regulatory functions (Yanhui et al. 2006) and we do not yet have enough data to infer the role of this gene in *S. appendiculatum*. Therefore, whether some upstream genetic changes trigger the downstream changes in pectin-related genes will have to be addressed in future studies. For instance, transcriptome analysis of additional developmental stages of male and female flowers could clarify how pectin regulation changes across flower development and the specific timing of divergent expression differences between male and female flowers. Regardless, with a genome-wide search for sex-specific sequences, in conjunction with gene expression analyses, we were able to detect both putative sex-determining regions and genes that may contribute to at least one of the two steps expected in the path from hermaphroditism to dioecy. These loci provide clear candidates for direct functional analysis in this system, especially for inaperturate pollen development phenotypes in female flowers.

The *S. appendiculatum* Genome Provides a Foundation for Addressing Repeated Transitions to Dioecy

Although the speciose genus *Solanum* contains fewer than 20 documented dioecious species, dioecy is estimated to have arisen independently at least 4 times (Anderson et al. 2015). Many of these transitions appear to involve common phenotypic features, most notably the development of inaperturate pollen in female individuals and dramatic reduction of the pistil in male flowers (Anderson et al. 2015). As such, this young genus (estimated ∼17 My old; Särkinen et al. 2013) offers a promising system in which to address the genomic features and genetic mechanisms of repeated, recent transitions to dioecy.

Solanum appendiculatum is among the most recently evolved dioecious angiosperms with sequenced genomes (<4 My; Echeverría-Londoño et al. 2020). The resources generated here provide a valuable framework for examining additional transitions to dioecy in the highly speciose genus, including a high-quality assembled genome, transcriptome characterization for annotation and gene expression analyses, and a set of candidate loci for directed exploration in parallel systems. Because most dioecious nightshades have similar sexual traits, including inaperturate pollen in the stamens of female flowers (Anderson et al. 2015), addressing the parallel origins of dioecy in this group can also address whether these transitions have followed convergent paths at genomic, genetic, and developmental levels. In conjunction with the *S. appendiculatum* genome, sequence data from other dioecious *Solanum* species can be used to dissect these parallel origins of sex determination in *Solanum*, including whether these exhibit similar genomic features (in terms of the number, size, and distribution of emerging sex-determination regions), draw on the same kinds of genomic/genetic changes (i.e., share orthologous sex-linked regions), and/or involve the same specific pathways and individual loci, including whether there is a general role for pectin-related loci in the early emergence of sexual differentiation. In this context, study of the genetic control of sex expression in species like S. *polygamum* and S. *conocarpum*—both of which bear anthers on female flowers, but that anthers are largely devoid of any pollen (Anderson et al. 2015)—could prove especially informative. Data from multiple recent, parallel systems will also be critical for testing the general predictions of theoretical models of the evolution of dioecy and assessing whether the complexity of genomic transitions that underpinning real empirical transitions matches well with these theoretical expectations.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

The project has been deposited in NCBI BioProject database (accession no. PRJNA561636). Assembled genome sequences were submitted to NCBI SUB9082420.

References

Adam H, Collin M, Richaud F, Beulé T, Cros D, Omoré A, Nodichao L, Nouy B, Tregear JW. 2011. Environmental regulation of sex determination in oil palm: current knowledge and insights from other species. *Ann Bot.* 108(8):1529–1537.

Akagi T, Henry IM, Kawai T, Comai L, Tao R. 2016. Epigenetic regulation of the sex determination gene MeGI in polyploid persimmon. *Plant Cell* 28(12):2905–2915.

Akagi T, Henry IM, Ohtani H, Morimoto T, Beppu K, Kataoka I, Tao R. 2018. A Y-encoded suppressor of feminization arose via lineage-specific duplication of a cytokinin response regulator in kiwifruit. *Plant Cell* 30(4):780–795.

Akagi T, Henry IM, Tao R, Comai L. 2014. A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. *Science* 346(6209):646–650.

Akagi T, Pilkington SM, Varkonyi-Gasic E, Henry IM, Sugano SS, Sonoda M, Firl A, McNeilage MA, Douglas MJ, Wang T. 2019. Two Y-chromosome-encoded genes determine sex in kiwifruit. Nat. Plants 5:801–809.

Anderson GJ. 1979. Dioecious *Solanum* species of hermaphroditic origin is an example of a broad convergence. *Nature* 282(5741):836–838.

Anderson GJ, Anderson MK, Patel N. 2015. The ecology, evolution, and biogeography of dioecy in the genus *Solanum*: with paradigms from the strong dioecy in *Solanum polygamum*, to the unsuspected and cryptic dioecy in *Solanum conocarpum*. *Am J Bot.* 102(3):471–486.

Anderson GJ, Levine DA. 1982. Three taxa constitute the sexes of a single dioecious species of *Solanum*. *Taxon* 31(4):667–672.

Anderson GJ, Symon DE. 1985. Extrafloral nectaries in *Solanum*. *Biotropica* 17(1):40–45.

Anderson GJ, Symon DE. 1989. Functional dioecy and andromonoecy in *Solanum*. *Evolution* 43(1):204–219.

Baldwin SJ, Husband BC. 2013. The association between polyploidy and clonal reproduction in diploid and tetraploid *Chamerion angustifolium*. *Mol Ecol.* 22(7):1806–1819.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B Methodol.* 57(1):289–300.

Bernardello LM, Anderson GJ. 1990. Karyotypic studies in *Solanum* section basarthrum (Solanaceae). *Am J Bot.* 77(3):420–431.

Bosch M, Hepler PK. 2005. Pectin methylesterases and pectin dynamics in pollen tubes. *Plant Cell* 17(12):3219–3226.

Bosch M, Hepler PK. 2006. Silencing of the tobacco pollen pectin methylesterase NtPPME1 results in retarded in vivo pollen tube growth. *Planta* 223(4):736–745.

Bräutigam K, Soolanayakanahally R, Champigny M, Mansfield S, Douglas C, Campbell MM, Cronk Q. 2017. Sexual epigenetics: gender-specific methylation of a gene in the sex determining region of *Populus balsamifera*. *Sci Rep.* 7:45388.

Campbell MS, Law MYee, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al. 2014. MAKER-P: a tool kit for the rapid creation, management, and quality control of plant genome annotations. *Plant Physiol.* 164(2):513–524.,

Caporali E, Carboni A, Galli M, Rossi G, Spada A, Longo GM. 1994. Development of male and female flower in *Asparagus officinalis*. Search for point of transition from hermaphroditic to unisexual developmental pathway. *Sexual Plant Reprod.* 7(4):239–249.

Charlesworth B, Charlesworth D. 1978. A model for the evolution of dioecy and gynodioecy. *Am Nat.* 112(988):975–997.

Charlesworth D. 2002. Plant sex determination and sex chromosomes. *Heredity* 88(2):94–101.

Charlesworth D. 2013. Plant sex chromosome evolution. *J Exp Bot.* 64(2):405–420.

Charlesworth D. 2015. Plant contributions to our understanding of sex chromosome evolution. *New Phytol.* 208(1):52–65.

Cheung AY, Wu H-M. 2008. Structural and signaling networks for the polar cell growth machinery in pollen tubes. *Annu Rev Plant Biol.* 59:547–572.

de Laat AM, Blaas J. 1984. Flow-cytometric characterization and sorting of plant chromosomes. *Theor Appl Genet.* 67(5):463–467.

Echeverría-Londoño S, Särkinen T, Fenton IS, Purvis A, Knapp S. 2020. Dynamism and context-dependency in diversification of the megadiverse plant genus *Solanum* (Solanaceae). *J Syst Evol.* 58(6):767–782.

Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet.* 8(9):689–698.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9(1):18.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.

Ezura K, Ji-Seong K, Mori K, Suzuki Y, Kuhara S, Ariizumi T, Ezura H. 2017. Genome-wide identification of pistil-specific genes expressed during fruit set initiation in tomato (*Solanum lycopersicum*). *PLoS One* 12(7):e0180003.

Gaffe J, Tieman DM, Handa AK. 1994. Pectin methylesterase isoforms in tomato (*Lycopersicon esculentum*) tissues (effects of expression of a pectin methylesterase antisense gene). *Plant Physiol.* 105(1):199–203.

Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220(4601):1049–1051.

Gou J-Y, Miller LM, Hou G, Yu X-H, Chen X-Y, Liu C-J. 2012. Acetylesterase-mediated deacetylation of pectin impairs cell elongation, pollen germination, and plant reproduction. *Plant Cell* 24(1):50–65.

Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26(9):1145–1151.

Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 30(8):1987–1997.

Han Y, Wessler SR. 2010. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 38(22):e199.

Harkess A, Mercati F, Shan HY, Sunseri F, Falavigna A, Leebens-Mack J. 2015. Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytol.* 207(3):883–892.

Harkess A, Zhou J, Xu C, Bowers JE, Van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, et al. 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat Commun.* 8(1):1279.,

Henry IM, Akagi T, Tao R, Comai L. 2018. One hundred ways to invent the sexes: theoretical and observed paths to dioecy in plants. *Annu Rev Plant Biol.* 69:553–575.

Jiang J, Zhang Z, Cao J. 2013. Pollen wall development: the associated enzymes and metabolic pathways. *Plant Biol.* 15(2):249–263.

Kazama Y, Ishii K, Aonuma W, Ikeda T, Kawamoto H, Koizumi A, Filatov DA, Chibalina M, Bergero R, Charlesworth D, et al. 2016. A new physical mapping approach refines the sex-determining gene positions on the *Silene latifolia* Y-chromosome. *Sci Rep.* 6:18917.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 12(4):357–360.

Kim S, Park J, Yeom SI, Kim YM, Seo E, Kim KT, Kim MS, Lee JM, Cheong K, Shin HS, et al. 2017. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* 18(1):210.

Knapp S, Bohs L, Nee M, Spooner DM. 2004. Solanaceae—a model for linking genomics with biodiversity. *Comp Funct Genomics.* 5(3):285–291.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15:356.

Levine DA, Anderson GJ. 1986. Evolution of dioecy in an American *Solanum*. In: D'Arcy WG, editor. Solanaceae: biology and systematics. New York: Columbia University Press. p. 264–273.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Li S-F, Zhang G-J, Yuan J-H, Deng C-L, Gao W-J. 2016. Repetitive sequences and epigenetic modification: inseparable partners play important roles in the evolution of plant sex chromosomes. *Planta* 243(5):1083–1095.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A. 2009. A transposon-induced epigenetic change leads to sex determination in melon. *Nature* 461(7267):1135–1138.

McDonnell AJ, Wetreich HB, Cantley JT, Jobson P, Martine CT. 2019. *Solanum plastisexum*, an enigmatic new bush tomato from the Australian Monsoon Tropics exhibiting breeding system fluidity. *PhytoKeys* 124:39–55.

Micheli F. 2001. Pectin methylesterases: cell wall enzymes with important roles in plant physiology. *Trends Plant Sci.* 6(9):414–419.

Mollet J-C, Leroux C, Dardelle F, Lehner A. 2013. Cell wall composition, biosynthesis and remodeling during pollen tube growth. *Plants* 2(1):107–147.

Murase K, Shigenobu S, Fujii S, Ueda K, Murata T, Sakamoto A, Wada Y, Yamaguchi K, Osakabe Y, Osakabe K, et al. 2017. MYB transcription factor gene involved in sex determination in *Asparagus officinalis*. *Genes Cells.* 22(1):115–123.

Nag A, Jack T. 2010. Sculpting the flower; the role of microRNAs in flower development. *Curr Top Dev Biol.* 91: 349–378.

Neal P, Anderson G. 2005. Are 'mating systems' 'breeding systems' of inconsistent and confusing terminology in plant reproductive biology? Or is it the other way around? *Plant Syst Evol.* 250(3-4):173–185.

Oh SA, Park KS, Twell D, Park SK. 2010. The SIDECAR POLLEN gene encodes a microspore-specific LOB/AS2 domain protein required for the correct timing and orientation of asymmetric cell division. *Plant J.* 64(5):839–850.

Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* 27(9):358–367.

Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of LTR retrotransposons. *Plant Physiol.* 176(2):1410–1422.

Renner SS. 2014. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot.* 101(10):1588–1596.

Rice WR. 1987. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41(4):911–914.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11(3):R25.

Särkinen T, Bohs L, Olmstead RG, Knapp S. 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evol Biol.* 13:214.

Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6(3):e17288.

Shi J, Cui M, Yang L, Kim Y-J, Zhang D. 2015. Genetic and biochemical mechanisms of pollen wall development. *Trends Plant Sci.* 20(11):741–753.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.

Smit AFA, Hubley R, Green P. 2013–2015. *RepeatMasker Open-4.0*. Available from: http://www.repeatmasker.org. Accessed April 04, 2021.

Song Y, Ma K, Ci D, Zhang Z, Zhang D. 2013. Sexual dimorphism floral microRNA profiling and target gene expression in andromonoecious poplar (*Populus tomentosa*). *PLoS One* 8(5):e62681.

Symon D. 1979. Sex Forms in Solanum and the role of pollen collecting insects. In: Hawkes JG, Lester RN, Skelding AD, editors. The Biology and Taxonomy of the Solanaceae. Linnean Society Symposium Series, 7. London: Academic Press. p. 385–396.

Symon DE. 1981. A revision of the genus *Solanum* in Australia. *J Adelaide Bot Gard.* 4:1–367.

Thomas GW, Hahn MW. 2019. Referee: reference assembly quality scores. *Genome Biol Evol.* 11(5):1483–1486.

Tomato Genome Consortium. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641.

Torres MF, Mathew LS, Ahmed I, Al-Azwani IK, Krueger R, Rivera-Nuñez D, Mohamoud YA, Clark AG, Suhre K, Malek JA. 2018. Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. *Nat Commun.* 9(1):3969.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33(14):2202–2204.

Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, Aryal R, VanBuren R, Murray JE, Zhang W, et al. 2012. Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci U S A.* 109(34):13710–13715.

Whalen MD, Costich DE. 1986. Andromonoecy in *Solanum*. In: D'Arcy WG, editor. Solanaceae: Biology and Systematics. New York: Columbia University Press. p. 284–302.

Wu H-C, Bulgakov VP, Jinn T-L. 2018. Pectin methylesterases: cell wall remodeling proteins are required for plant response to heat stress. *Front Plant Sci.* 9:1612.

Wu M, Kostyun J, Moyle L. 2019. Genome sequence of Jaltomata addresses rapid reproductive trait evolution and enhances comparative genomics in the hyper-diverse Solanaceae. *Genome Biol Evol.* 11(2):335–349.

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, et al. 2006. The MYB transcription factor superfamily of *Arabidopsis*: expression analysis and phylogenetic comparison with the rice MYB family. *Plant Mol Biol.* 60(1):107–124.

Zavada MS, Anderson GJ. 1997. The wall and aperture development of pollen from dioecious *Solanum appendiculatum*: what is inaperturate pollen? *Grana* 36(3):129–134.

Zhang GY, Feng J, Wu J, Wang XW. 2010. BoPMEI1, a pollen-specific pectin methylesterase inhibitor, has an essential role in pollen tube growth. *Planta* 231(6):1323–1334.

Zhu J, Chen H, Li H, Gao JF, Jiang H, Wang C, Guan YF, Yang ZN. 2008. Defective in Tapetal development and function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *Plant J.* 55(2):266–277.

Zhu J, Zhang G, Chang Y, Li X, Yang J, Huang X, Yu Q, Chen H, Wu T, Yang Z. 2010. AtMYB103 is a crucial regulator of several pathways affecting *Arabidopsis* anther development. *Sci China Life Sci.* 53(9):1112–1122.

Zimin AV, Marcais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29(21):2669–2677.

Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the mega-reads algorithm. *Genome Res.* 27(5):787–2677.