We present a draft genome assembly for the tropical liverwort, *Marchantia inflexa*, which adds to a growing body of genomic resources for bryophytes and provides an important perspective on the evolution and diversification of land plants. We specifically address questions related to sex chromosome evolution, sexual dimorphisms, and the genomic underpinnings of dehydration tolerance. This assembly leveraged the recently published genome of related liverwort, *M. polymorpha*, to improve scaffolding and annotation, aid in the identification of sex-linked sequences, and quantify patterns of sequence differentiation within *Marchantia*. We find that genes on sex chromosomes are under greater diversifying selection than autosomal and organellar genes. Interestingly, this is driven primarily by divergence of male-specific genes, while divergence of other sex-linked genes is similar to autosomal genes. Through analysis of sex-specific read coverage, we identify and validate genetic sex markers for *M. inflexa*, which will enable diagnosis of sex for non-reproductive individuals. To investigate dehydration tolerance, we capitalized on a difference between genetic lines, which allowed us to identify multiple dehydration associated genes two of which were sex-linked, suggesting that dehydration tolerance may be impacted by sex-specific genes.
individuals), has eight autosomes and one female (U) or male (V) sex chromosome, and reproduces sexually by spores, or asexually by fragmentation or the formation of gemmae (specialized asexual propagules). *Marchantia inflexa* typically grows on rock and soil surfaces along stream banks in tropical forests, but can also colonize more exposed and disturbed sites along roads. Vegetative growth produces a dichotomously branching thallus mat with dorsiventral organization, and the haploid gametophyte is the dominant life phase. *Marchantia inflexa* is a useful model to investigate sexual dimorphisms, population sex ratios, and stress tolerance because prior work has established that *M. inflexa* exhibits a considerable degree of sexual dimorphism in sex ratio, variable population sex ratios, and fluctuating stress tolerance.

Bryophytes harbor a high proportion of dioecious species. Nearly half of all extant mosses, and approximately two-thirds of liverworts are dioecious. In many bryophytes (including *M. inflexa*) the reported sex ratio is often female biased, and in *M. inflexa*, this may be related to females' superior ability to recover from drying events, faster growth rate, or the increased establishment of female gemmae. However, true population sex ratios are largely unknown, except for the few cases where genetic sex markers have been developed and utilized. Typical methods for assessing sex ratios depend on counting the number males and females with visible sex organs and using this to infer the underlying population sex ratio. However, this approach fails to account for plants not currently displaying sex organs and assumes that the sex ratio of vegetative plants is equivalent to that of plants with sex organs. This assumption may not hold true in natural settings. In fact, for both *M. polymorpha* and *M. inflexa* (where sex organ development can be artificially induced) the timing of reproductive development is sex-specific and some individuals never produce sex organs (unpublished data).

The reproductive biology of bryophytes (with the haploid gametophyte being the dominant life stage) provides a unique perspective on the evolution of sex-linked genes, as the female (U) and male (V) sex chromosomes are present at the same copy number (1N) as autosomal chromosomes for the majority of the organism’s life cycle and are subject to haploid selection. Sex chromosome evolution in diploid dominant systems has received considerable research attention. However, less is known about the forces shaping sex chromosomes in haploid dominant systems, and the ramifications of haploid selection on sex chromosomes may have unique consequences. For example, exposure to haploid selection should reduce the prevalence of deleterious mutations and could allow beneficial mutations to spread more rapidly. However, lack of recombination on UV sex chromosomes could lead to degeneration on UV chromosomes, as has been observed in XY andZW chromosomes. Further, the smaller effective population size of sex chromosomes relative to autosomes may increase the impact of genetic drift, further influencing adaptive evolution of sex-specific genes. The extent to which these forces shape sex chromosome evolution in haploid dominant systems is not well understood, but the numerous dioecious bryophyte taxa provide novel opportunities to test related questions.

Stresses caused by environmental fluctuations are accentuated in plants due to their sessile nature. Consequently, numerous tolerance mechanisms have evolved to combat environmental pressures, many of which have potential translational utility. Some of these stress tolerance traits, such as embryo retention (allowing for the development and dispersal of desiccation tolerant spores), UV radiation, desiccation, heat, and freezing tolerance may have facilitated the transition from aquatic to terrestrial environments by early plants. Many extant bryophytes retain these early stress tolerance mechanisms, allowing them to occupy marginal niches characterized by nutrient poor substrates, toxic concentrations of metals, variable light and moisture levels. Consequently, bryophytes are particularly informative with respect to understanding the evolutionary history and physiological strategies of stress tolerance.

Desiccation tolerance (DT) in particular, has important translational utility. A number of studies have described the genomes of DT plants, and the amassed data provide a strong foundation on which to construct our understanding of DT. These studies have demonstrated that DT is a complex multigenic trait, and that there are multiple means of achieving DT. The genetic basis of DT, although not entirely described, may derive from regulatory differences in gene expression pathways, increased copy number of anahydrobiosis related genes, or differences in the structural organization of these genes. However, more studies are needed to resolve the specifics of DT mechanisms, and should include work on species spanning a wide phylogenetic range and degree of tolerance levels (such as the intermediate trait of dehydration tolerance (DhT) also dehydration tolerant). *Marchantia inflexa* is DhT, which provides an important opportunity to enhance our understanding of the evolution of this intermediate trait.

Growing genomic resources for bryophytes provide novel opportunities to conduct comparative studies within these lineages, which are particularly well suited to addressing questions related to sex chromosome evolution, sex differences, and stress tolerance adaptations. Here, we aimed to characterize patterns of sequence variation within these lineages, which are particularly well suited to addressing questions related to sex chromosome evolution, sex differences, and stress tolerance adaptations. We, therefore, aimed to characterize patterns of sequence variation within these lineages, which are particularly well suited to addressing questions related to sex chromosome evolution, sex differences, and stress tolerance adaptations.
DhT in *M. inflexa* females. The remaining DhT genes appear to be located on autosomes and are expressed at similar levels in both sexes, which may contribute to the changing patterns of DhT that have been observed.

**Results**

**Genome assembly and annotation.** Whole-genome sequencing of *M. inflexa*, yielded 127,147,280 male reads and 133,660,960 female reads (after quality filtering). The combined male and female k-mer distribution indicated a coverage of ~24x, but showed a large quantity of unique and low abundance k-mers, suggestive of contaminating organisms (see Supplementary Fig. S1). In our efforts to characterize the source of these low abundance k-mers, we detected a diverse community of microbes, consistent with recent descriptions of *M. inflexa* microbial associations. After removal of putative microbial sequences, we assembled the remaining sequence reads to generate the draft assembly *M. inflexa* v1.1. The resulting scaffolds were assigned to super-scaffolds by alignment with the *M. polymorpha* reference genome, allowing us to coalesce the assembly into 300 super-scaffolds. In total 7,747 *M. inflexa* scaffolds covering a total length of 81,634,927 bp were successfully mapped to the *M. polymorpha* genome. Unmapped *M. inflexa* scaffolds were appended to the super-scaffold assembly. The resulting assembly consists of 41,556 scaffolds, covering a total of 208,839,958 bp, with an N50 of 11,144 bp and the longest scaffold length of 2,829,880 bp. This Whole Genome Shotgun project (*M. inflexa* v1.1) has been deposited at DDBJ/ENA/GenBank under the accession QLSQ00000000. The version described in this paper is version QLSQ01000000.

Assessment of assembly completeness (performed with BUSCO) indicated that 54.4% (783) of the 1,440 presumptively universal single-copy orthologs from the plant set of OrthoDB v9 were present in the *M. inflexa* genome assembly. Another 3.5% (51) orthologs were present, but fragmented. In comparison, a parallel assessment of the *M. polymorpha* v3.1 assembly, found that 60.2% (867) of these same genes were complete, and 2.9% (42) were fragmented in *M. polymorpha*. Both of these estimates are rather low, suggesting that there may be inherent limitations associated with BUSCO as has been observed for other deeply diverged lineages. Still, we find these assessments to be informative in a comparative context within Marchantia.

Assembly of *M. inflexa* plastids generated nearly complete mitochondrial and chloroplast sequences (Fig. 1). The mitochondria of *M. inflexa* is 190,056 bp and the chloroplast is 122,620 bp. The complete mitochondrial and chloroplast sequences are available at FigShare (https://doi.org/10.6084/m9.figshare.6639209.v1).

**Annotation.** Gene annotation of the *M. inflexa* draft genome utilized de novo gene finding in combination with the lift-over of all *M. polymorpha* annotations for orthologous genes. Lift-over annotations from *M. polymorpha* v3.1 assembly, found that 60.2% (867) of these same genes were complete, and 2.9% (42) were fragmented in *M. polymorpha*. Both of these estimates are rather low, suggesting that there may be inherent limitations associated with BUSCO as has been observed for other deeply diverged lineages. Still, we find these assessments to be informative in a comparative context within Marchantia.
M. inflexa, P. patens and A. thaliana, reflecting the estimated divergence times among these species (divergence time between M. inflexa and M. polymorpha is 68–126 MYA; for M. inflexa and P. patens it is 425–557 MYA; and for M. inflexa and A. thaliana it is 481–584 MYA).  

Sequence similarity between M. inflexa and M. polymorpha. To investigate genome evolution within Marchantia we measured sequence divergence between M. inflexa and M. polymorpha. Initially, we compared nucleotide differentiation among coding sequences (CDS), introns, and intergenic regions to estimate general patterns of divergence between lineages (Fig. 2). Comparison of orthologous CDS, introns, and intergenic sequences, revealed that (not surprisingly) intergenic sequences were the least conserved (64.5% ± 0.009%), introns were intermediate (81.8% ± 0.008%), and CDS were the most conserved (82.4% ± 0.001%) (Fig. 2). There was a significant effect of sequence type on %ID (F2,40000 = 39756, p < 0.0001). Patterns of sequence divergence between M. inflexa and M. polymorpha fit general expectations that CDS should exhibit higher sequence similarity compared to introns and intergenic sequences. That being said, we observed surprisingly high sequence conservation among some introns, which we speculate is related to the relatively short length of M. inflexa introns, in which functional elements (such as splice sites) may be preferentially retained.

In order to assess variation in substitution rates across coding sequences, we computed the ratio of non-synonymous to synonymous mutations (dN/dS) for all orthologous CDS of M. inflexa and M. polymorpha. The resulting dN/dS values were log transformed to improve normality for statistical testing. Initially, we tested for evidence of contrasting selective pressures among autosomal, sex-linked, and organellar genes. Notably, sex linked genes and autosomes are present at 1 N, whereas copy number of the chloroplast and mitochondria is variable. We computed mean dN/dS and standard error for autosomal genes (0.24 ± 0.01 (n = 4,900)), for sex-linked genes (0.48 ± 0.13 (n = 53)), and for organellar genes (0.14 ± 0.03 (n = 116)). We detected significant differences among groups (F2,4862 = 18.54, p < 0.001). Targeted contrasts revealed significant differences among sex-linked and autosomal genes (t1 = −2.38, p = 0.018) and among organellar and autosomal genes (t1 = 5.58, p = 2.6e−8). Subsequently, we tested for differences among more specific gene types; subdividing sex-linked genes into male-specific, female-specific, and male and female alleles of genes with both U and V copies. Organellar genes were subdivided into mitochondrial and chloroplast genes. Mean dN/dS of male-specific genes was 0.63 ± 0.23 (n = 23), of female-specific was 0.20 ± 0.09 (n = 7), of male alleles was 0.24 ± 0.12 (n = 11), and of female alleles was 0.56 ± 0.34 (n = 12). Mean dN/dS of chloroplast genes was 0.03 ± 0.01 (n = 74) and of mitochondria was 0.34 ± 0.06 (n = 42). There was an overall effect of gene type on dN/dS (F2,4862 = 20.10, p < 0.001). Targeted contrasts revealed significant differences between autosomal genes and male-specific genes (t1 = −2.88, p = 0.004), chloroplast genes (t1 = 10.39, p = 5e−25), and mitochondrial genes (t1 = −4.40, p = 1.1e−5) (Fig. 3).

These analyses reveal several genes and pathways that may be under diversifying selection (dN/dS > 1) in M. inflexa and M. polymorpha. Sex-linked genes with dN/dS > 1 included the female allele of CCR4-NOT transcription related complex protein (Mapoly0018s0021.1), the male bHLH-MYC2 transcription factor in the involved the jasmonate signaling pathway (MapolyY_B0018.1), a male-specific phosphatidylinositol-4,5-bisphosphate 3-kinase (MapolyY_A0049.1), two male-specific genes of unknown function (MapolyY_B0032.1 and MapolyY_B0003.1). No chloroplast genes in our analyses had dN/dS > 1, but three mitochondrial open reading frames (orf 84, orf 69, rpl10) had dN/dS > 1. Of the 243 autosomal genes with dN/dS > 1, 51 had identifiable homologs in the Uniprot database. GO analyses of these genes revealed that many were associated with the cellular components.
intracellular, cytoplasm, and membrane, the molecular function catalytic activity (followed closely by hydrolase activity and transferase activity), and the biological processes of metabolic process and cellular process. A complete list of genes with dN/dS > 1 and associated protein names can be found as Supplementary Table S1.

Sex marker identification. We identified 4,468 regions (covering 2,234,000 bp) in the *M. inflexa* genome assembly with substantial differences in copy number among genetic lines through coverage analysis with DifCover (https://github.com/timnat/DifCover)65 (see Supplementary Fig. S2). Of these, 89 were found on scaffolds also containing a predicted protein, 31 of which could be assigned to an identifiable homolog across *M. polymorpha*, *P. patens*, *A. thaliana*, and refseq databases. From this set, we identified five putatively male- and three female-specific sequences that were also orthologous to sequences on the U and V chromosomes in *M. polymorpha*. These candidate sex markers were analyzed by PCR in nine males and nine females to verify their fidelity, leading to the validation of one positive marker for each sex (Fig. 4). Other candidate sex markers exhibited non-specific amplification and were therefore discarded. Plants used for validation were originally collected from five distinct populations, suggesting that the markers are robust to genotypic variation. Primer sequences of the validated male and female sex markers are listed in Table 1.

Dehydration tolerance. To address specific hypotheses on DiT we probed *M. inflexa* and *M. polymorpha* annotated proteins for orthologs to a list of 195 DT genes (compiled from publicly available mRNA sequences of genes expressed under water stress in model DT plants). See Supplementary Table S2 for the accession numbers, species, and studies from which genes were compiled. Of this set of DT genes, 112 had identifiable homologs in *M. inflexa* and 141 had identifiable homologs in *M. polymorpha*. Our analyses of dN/dS captured 38 of these DT orthologs, one of which (a putative aldehyde dehydrogenase (Mapoly0030s0099.1)) had a dN/dS value > 1. The
function of diversification in this gene is unclear, given the lack of evidence for any difference in DhT between these two Marchantia species.

Prior studies showed that the male and female M. inflexa genotypes used for genome assembly have reproducible differences in DhT. Consequently, we aimed to identify DT genes with substantial coverage differences among these two genotypes, presuming that they may impact relative differences in DhT. Of the 112 DT genes detected in M. inflexa, most had standardized coverage ratios of ~1. However, six genes had considerably higher coverage (log2 fold change > 4) in the highly tolerant female and one had higher coverage in the less tolerant male (Table 2). Specific genes with higher coverage in the tolerant female genotype include a calcium dependent protein kinase (CDPK), glucose related protein 94 (GRP94), the aldehyde dehydrogenase (ALDH) (also identified in dN/dS analyses above), heat shock proteins 70 and 101 (HSP101, HSP70), and superoxide dismutase (SOD). The sole DT gene with higher coverage in the less tolerant male genotype is a heat shock factor 1 (HSF 1). Of the DT genes with coverage difference among genotypes, one (CDPK) was assigned to the putative U chromosome, and one (HSF 1) was assigned to the putative V chromosome.

To verify that candidate DT associated genes were expressed during dehydration and to test for sex-specific patterns of expression, we conducted qPCR validation of seven candidate dehydration associated genes. We quantified expression for each gene in three males and three females under both hydrated and dehydrated conditions by qPCR. These analyses revealed that all candidate genes were expressed in M. inflexa under both hydrated and dehydrated conditions (Fig. 5). Analysis of changes in relative expression of DhT associated genes during dehydration identified an overall increase in expression during dehydration (F1,1 = 5.70, p = 0.019), differences among genes (F0.6 = 9.03, p < 0.001) and a significant interaction between gene and hydration state (F0.6,3 = 2.41, p = 0.0351). The interaction effect was driven primarily by an increase in HSP70 expression during dehydration (other genes did not show significant changes in expression during dehydration). Interestingly, the two candidate genes that were putatively sex-linked (female-specific CDPK and male-specific HSF 1) exhibited sex-specific expression, but autosomal candidate DhT genes (GRP94, ALDH, HSP101, HSP70, and SOD) were expressed at similar levels in males and females suggesting copies are present in both sexes. These autosomal genes were present at different copy number in the male and female used for genome assembly, but in the larger panel assayed by qPCR this did not translate to differences in relative expression among the sexes.

### Discussion

Our assembly of the M. inflexa genome represents a new resource for comparative studies among land plants. We capitalized on the recently published genome of related liverwort M. polymorpha to improve scaffolding of our assembly, estimate divergence rates among specific sequence types, and to identify sex-linked sequences that were leveraged to generate male and female genetic sex markers for M. inflexa. Our analyses identify several genes on the autosomes, organelles, and sex chromosomes that show strong signatures of recent diversifying selection in Marchantia. Additionally, we identified multiple genes possibly underlying an observed genotype difference in DhT in M. inflexa, which point towards a complex mechanism of heightened DhT. We detected differences in copy number of DhT genes across multiple loci, two of which were putatively sex-linked. Evidence of sex-linked genes underlying differences in DhT is intriguing, as prior studies indicate complex patterns of sexual dimorphism in DhT in M. inflexa.

Analyses of dN/dS ratios for genes on autosomes, sex chromosomes, and organelles in M. inflexa and M. polymorpha showed evidence of increased diversification of sex-linked genes relative to autosomal genes, and conservation of organelar genes (particularly the chloroplast) relative to autosomes. UV sex determination systems are expected to differ from diploid dominant (XY and ZW) sex determination systems in multiple ways, due primarily to haploid selection and theoretical studies on sex chromosome evolution.

![Table 1. Primer sequences for validated male and female genetic sex markers for Marchantia inflexa.](image)

| Marker          | Left primer sequence | Right primer sequence |
|-----------------|----------------------|-----------------------|
| Male marker 98683 | CGTTGTATCCTCTTCTCAG  | AGCTTTGTCAGAAATAGTCAG |
| Female marker 42793 | GTCCATGTCGAAAGCGTA  | CCTCTTCGATACAGTCAGT |

![Table 2. DhT proteins with coverage differences among the sexes in Marchantia inflexa. We designated a log2 fold change > 4 as the cutoff to define a significant coverage difference. Negative log2 coverage ratios indicate higher female coverage relative to males, whereas positive log2 coverage ratios indicate higher male coverage.](image)

| Protein                      | Log2 (male coverage/female coverage) |
|------------------------------|--------------------------------------|
| Heat shock factor 1 (HSF 1)  | 14.3397                              |
| Calcium dependent protein kinase (CDPK) | −9.7936                           |
| Glucose related protein 94 (GRP94) | −16.4576                           |
| Aldehyde dehydrogenase (ALDH) | −4.2169                              |
| Heat shock protein 70 (HSP101) | −7.318                               |
| Heat shock protein 101 (HSP70) | −16.4576                             |
| Superoxide dismutase (SOD)    | −4.995                               |
Complex (dN/dS) due to differences in sex function in other systems. Notably, the female-allele of CCR4-NOT transcription related mechanism for maintaining species boundaries. Developed DhT in M. inflexa. Of these genes were putatively sex-linked, suggesting that there may be sex-specific components to DhT in M. inflexa. Interestingly two genes (Fig. 3) were putatively sex-linked (female-specific CDPK and male-specific HSF1) were identified several such genes, which we speculated contribute to differences in DhT among genotypes. Error bars are standard error of the mean.

Figure 5: Relative expression of candidate dehydration associated genes in three male and three female plants under hydrated and dehydrated conditions. Gene expression estimates were standardized using an internal control gene (actin). The expression of each candidate gene relative to actin was computed to estimate relative abundance. Absence of a bar (as seen for HSF 1 and CDPK) indicates that there was no amplification of that gene in the corresponding sample. Error bars are standard error of the mean.

In our efforts to identify genes underlying DhT we capitalized on a previously identified genotype difference in DhT in M. inflexa, targeting genes that exhibited substantial coverage differences among genotypes. We identified several such genes, which we speculated contribute to differences in DhT among genotypes. Interestingly two of these genes were putatively sex-linked, suggesting that there may be sex-specific components to DhT in M. inflexa. Importantly, qPCR validation of seven candidate dehydration associated genes in additional male and female plants revealed persistent expression of these genes under hydrated and dehydrated conditions, supporting a possible role for these genes in DhT. The unique biology of bryophytes makes them prone to rapid drying and thus, many genes involved in DhT are expressed prior to dehydration in bryophytes. Various patterns of expression have been observed in DhT and DT bryophytes, including constitutive expression, expression induced during drying, and the sequestration of mRNAs in ribonucleoprotein particles. Thus, there is no absolute predication to how a given gene will be expressed during dehydration. Our qPCR analyses suggest that even in this small set of candidate DhT genes, multiple patterns of expression are evident in M. inflexa. The majority of tested genes were expressed constitutively, but one (HSP70) showed a significant increase in expression during dehydration. Importantly, the two genes that were putatively sex-linked (female-specific CDPK and male-specific HSF1) were expressed in a strictly sex-specific pattern. Interestingly CDPKs have been recognized as important hubs in plant stress signaling pathways with highly conserved structure, which provides a possible explanation for elevated DhT in females. Other candidate DhT genes were expressed at similar levels across the sexes, suggesting that
Despite possible involvement in DhT these genes are unlikely to drive sex-related differences in DhT within *M. inflexa*. Quantification of candidate DhT gene expression by qPCR further reinforces gene annotations in demonstrating that predicted *M. inflexa* genes are expressed in the expected tissues and individuals.

In summary, the draft genome for *M. inflexa* adds to a growing body of genomic resources for land plants, which will enable investigation of early plant evolution and physiology. We leveraged this assembly to identify genes under diversifying selection in *Marchantia*, to develop genetic sex markers, and to target genes contributing to DhT. Our analyses comprise one of the few empirical studies on haploid sex chromosome evolution and suggest that several sex-linked genes (particularly male-specific (V) genes) have undergone rapid diversification in *Marchantia*. We identified multiple sex-specific sequences, which were used to develop genetic sex markers and identify genes underlying differences in DhT of *M. inflexa*. We find evidence that DhT in *M. inflexa* is likely impacted by the constitutive expression of select DhT genes, and that sex differences in DhT may be impacted by sex-linked DhT genes.

**Methods**

**Plant growth, DNA extraction, and sequencing.** Plant specimens for genome sequencing were collected from East Turure stream (10°41’04”N 61°09’39”W) on the island of Trinidad, Republic of Trinidad and Tobago in 2009. Voucher specimens are deposited at the Missouri Botanical Garden (St. Louis, MO, USA, specimen numbers M092113 and M092115) and at the National Herbarium of the Republic of Trinidad and Tobago (St. Augustine, Trinidad, specimen number TRIN34616, D. N. McLetchie, collector). Vegetative tissue was transported to Lexington, Kentucky, USA and thirty-six clones (generated though vegetative propagation) of one male and one female genotype were planted on steam-sterilized soil and maintained in a randomized layout in a climate-controlled greenhouse. Plants were watered daily with distilled water and kept under shade cloth to mimic field conditions. Vegetative tissue (growing annually with no soil contact) was collected from male and female plants after ~5 years in greenhouse conditions. Prior to DNA extraction, thalli were washed in distilled water three times to remove surface contamination. DNA was extracted following a CTAB extraction protocol modified from Doyle73. Sequencing libraries were constructed with 300 base pair (bp) inserts and whole genome sequencing was conducted on an Illumina HiSeq2000 for 100 bp paired end (PE) reads at the Beatty Biodiversity Research Centre, University of British Columbia.

**Genome assembly and annotation.** Sequence read quality was assessed with fastQC version 384, and filtered with Trimomatic version 0.3373. Male and female reads were combined to increase coverage and k-mer plots were generated with DSK version 1.1 (see Supplementary Fig. S1)79. Assembly was carried out using SOAP de novo version 2.04-r24077 with a k-mer length of 31. Reads shorter than 100 bp were not included, alignments of less than 32 bp were not considered reliable, and k-mers observed nine or fewer times were excluded from the assembly.

Following initial assembly, we plotted the length and GC content of each scaffold in JMP®, Version 12 (SAS Institute Inc.). The plot revealed two distinct clusters of well-assembled (long) scaffolds: one with a mean GC content of ~65% and one with a mean GC content of ~45% (see Supplementary Fig. S3). Consequently, we probed each distinct GC cluster to identify the taxonomic source of the contributing sequence reads by aligning the 100 longest scaffolds of each GC cluster to NCBI’s refseq database78 using TBLASTX79. Taxonomic classification of the resulting alignments with Megan version 480 revealed that scaffolds with high GC content were derived from a diverse microbial community, whereas scaffolds with low GC content were derived exclusively from plant material (see Supplementary Fig. S4). Notably, other members of *Marchantia* have a GC content of ~45%85, providing additional support for the assumption that low GC content reads were derived from *M. inflexa*. Consequently, we filtered the raw sequence data to remove all reads with a GC content >55%. The remaining reads (although likely not entirely contamination free) represent a data set enriched for *M. inflexa* genomic information. Using only reads with a GC content <55%, we reassembled the sequence data with the same parameters as above.

The resulting scaffolds were aligned to the *M. polymorpha* reference genome2 with BLASTN, allowing us to organize *M. inflexa* scaffolds with Chromosomer version 1.336, which leverages pairwise sequence alignments and local synteny to assign orthologous regions to super-scaffolds. All *M. inflexa* scaffolds not mapped to *M. polymorpha* were appended to the super-scaffold assembly, and all contigs under 1000 bp were removed. Assembly statistics were computed using assemblathon_stats_2.pl script (https://github.com/lexnderbragt/denovo-assembly-tutorial/blob/master/scripts/assemblathon_stats_2.pl). To estimate assembly completeness, we quantified the percentage of Universal Single-copy Orthologs from the plant set of OrthoDB v9 of BUSCO v364 present in the *M. inflexa* genome assembly. We conducted a parallel assessment of the *M. polymorpha* genome assembly.

To assemble the mitochondrial and chloroplast genomes of *M. inflexa* raw reads were trimmed with Trimomatic version 0.3373 and error corrected using the ErrorCorrectReads module of Allpaths-LG version 501356. These were aligned to the *M. polymorpha* reference plastid and mitochondrial sequences (GI 1146667384 and GI 1146709711, respectively) with BWA mem version 0.7.12-r103985. Reads with alignments, plus their mates, were extracted and partitioned. Each partition was assembled separately with Ray de novo version 2.3.1891 for four iterations of gene finding. The resulting predicted proteins were aligned to the *M. polymorpha*
Sequence similarity between *M. inflexa* and *M. polymorpha.* To enable comparison of orthologous sequences, we aligned the entire *M. inflexa* assembly to the *M. polymorpha* assembly v3.1 with LASTZ version 1.04, and extracted orthologous CDS, introns, and intergenic sequences from both assemblies using a combination of BEDTools version 2.19.184 and BEDOPS version 2.4.3595. To explicitly test for differences in nucleotide differentiation among these sequence types, CDS, introns, and intergenic sequences were realigned to one another using LASTZ version 1.04. The resulting mean % identity (%ID) for each sequence type was computed and differences among sequence types were tested for significance with a mixed effects linear model in JMP®, Version 12 (SAS Institute Inc.). The fixed effect of sequence type on %ID was tested (sequence length was included in the model as a random effect).

To investigate patterns of gene divergence we computed dN/dS for all orthologous CDS in *M. inflexa* and *M. polymorpha.* Initially, we extracted the complete CDS and translated amino acid sequence for all orthologous genes using gffread (https://github.com/gpete/gffread). Orthologous translated CDS were aligned with Clustal Omega version 1.2.496, and codon aware DNA alignments were defined using PAL2NAL version 14.187, during which all gaps and internal stop codons were removed. Next, dN/dS ratios for each ortholog were calculated with the yn00 function of PAML version 4.998, which computes dN/dS using pairwise comparisons accounting for both transition/transversion bias and base/codon frequency bias99. Following filtering conventions100, cases in which dS = 0, dN > 2, and dN/dS > 10 were removed from the output, and dN/dS values were log transformed to satisfy assumptions of normality for statistical testing. Differences among dN/dS values were tested for significance using a mixed effects linear model in JMP®, Version 12 (SAS Institute Inc.). Initially, the fixed effect of gene type (autosomal, sex-linked, or organellar) on dN/dS was tested (scaffold ID was included in the model as a random effect). Post hoc comparisons among gene types were made using orthogonal contrasts to explicitly compare autosomal genes to sex and organellar genes. Subsequently, we made more detailed comparisons among specific gene types with a mixed effects linear model testing the fixed effect of specific gene type (autosomal, male-specific, female-specific, male-allele, female-allele, mitochondria, and chloroplast) on dN/dS (scaffold ID was included in the model as a random effect). Again, post hoc comparisons were made using orthogonal contrasts to specifically compare each gene type to the autosomal genes. Finally, all individual genes with dN/dS values > 1 were identified, and gene ontology (GO) terms were defined with the GORetriever tool and summarized with the GOSlimViewer tool available at AgBase (http://agbase.msstate.edu/cgi-bin/tools.pl)101.

Sex marker identification. Read coverage was computed using DifCover (https://github.com/timnat/DifCover)102 to identify regions of the genome unique to these male and female genotypes. Briefly, we determined the genotype-specific coverage by mapping male and female sequence reads back to the draft assembly with Bowtie2103. Coverage was calculated for 500 bp windows with BEDtools version 2.19.184, and DifCover was used to explicitly test for differences in nucleotide differentiation among these sequence types, CDS, introns, and intergenic sequences were realigned to one another using LASTZ version 1.04. The resulting mean % identity (%ID) for each sequence type was computed and differences among sequence types were tested for significance with a mixed effects linear model in JMP®, Version 12 (SAS Institute Inc.). The fixed effect of sequence type on %ID was tested (sequence length was included in the model as a random effect). Post hoc comparisons among sequence types were made using orthogonal contrasts to explicitly compare each gene type to the autosomal genes. Finally, all individual genes with dN/dS values > 1 were identified, and gene ontology (GO) terms were defined with the GORetriever tool and summarized with the GOSlimViewer tool available at AgBase (http://agbase.msstate.edu/cgi-bin/tools.pl)101.

Dehydration tolerance. To address specific hypotheses on DhT we probed *M. inflexa* and *M. polymorpha* annotated proteins for orthologs to a list of 195 DT genes (compiled from publicly available mRNA sequences of genes expressed under water stress in model DT plants (see Supplementary Table S2)). DT orthologs were identified using BLASTX104, and the single best hits for each *M. inflexa* and *M. polymorpha* sequence were determined based on bitscore. Presuming that some DT genes would be multi-copy in *M. inflexa*, we calculated the genotype-specific coverage of each DhT ortholog using DifCover102 with the aim of detecting genes contributing to the observed genotype difference in DhT in *M. inflexa*. We targeted sequences corresponding to DT genes showing log, coverage ratios > 4 as potential contributors to the observed difference in DhT. Finally, we computed dN/dS ratios for all DhT genes and identified those exhibiting signs of diversifying selection (dN/dS > 1).

To verify that candidate dehydration associated genes were expressed during dehydration and to test for sex-specific patterns of expression, we conducted qPCR validation of seven candidate genes. To do so, we selected three male and three female plants from East Turure stream (see Supplementary Table S3). Plants were cultivated in a common garden for ~eight years prior to being subjected to dehydration treatment as described in105. Briefly, thallus tissues (~5 mm × 7 mm) were harvested from greenhouse cultivated plants, fully hydrated for 24 hours, and placed into desiccation chambers with an internal relative humidity (RH) of 75%. Thallus tips were randomized in Petri dishes within in the desiccation chamber. Each desiccation chamber contained 18 thallus tips (three tips from each of the six genotypes). Air circulation was maintained by inserting a small fan in the chamber, and RH was verified with a RH sensor integrated into HOBO™ humidity sensor attached to a data
logger (Onset Computer Corporation, Bourne, MA, USA). The chamber was maintained at 14 °C for 22 hours. Dehydration assays were conducted at designated times of day to reduce off target variation due to fluctuations in light, temperature, and circadian rhythms.

Plants were sampled at two time points during the dehydration assay: initial hydrated conditions and dehydrated conditions (22 hours). At each time point, tissues were removed from the desiccation chamber and immediately flash frozen in liquid nitrogen to prevent further transcriptional changes. RNA was extracted from samples using the Triazol® Reagent according to the manufacturer’s instructions, mRNA was isolated through poly(A) enrichment with the NEBNext® Poly(A) mRNA Magnetic Isolation Module, and cDNA synthesis was conducted with SMARTScribe™ Reverse Transcriptase using random primers. Biological replicates (genotypes) were processed independently and randomized for all downstream processing and analyses to minimize possible batch effects.

We measured gene expression for seven dehydration associated genes (GRP94, ALDH, HSP101, HSP70, SOD, female-specific CDPK, and male-specific HSF 1) by qPCR. The housekeeping gene actin was included as an internal control. Primer pairs for each gene were designed using Primer3® and are listed Supplementary Table S4. qPCR reactions were carried out in 96 well plates on a Roche LightCycler® 96. The reaction mix consisted of 1ul cDNA, 0.5ul each of the forward and reverse primers (10μM), 10ul Luna® Universal qPCR Master Mix, and 8ul PCR-grade H2O, for a final reaction volume of 20ul. A template free control was included for each primer pair. The amplification program consisted of a 60 second preincubation at 95 °C, followed by 45 cycles of 95 °C for 5 seconds, 60 °C for 15 seconds, and 72 °C for 10 seconds. Following amplification, high resolution melting analyses was conducted, via one cycle of 95 °C for 60 seconds, 40 °C for 60 seconds, 65 °C for 1 second, and 97 °C for 1 second.

Data were analyzed using the Roche LightCycler® 96 Software Version 1.1. Initially, melting curve analysis was conducted to confirm product quality, and any samples with abnormal profiles were removed from the analyses. Raw Cq (quantitation cycle) values were used to estimate cDNA concentration for the seven candidate dehydration associated genes and housekeeping gene actin (used as an internal control) (see Supplementary Fig. S6). Cq indicates the first cycle at which fluorescence could be detected (smaller Cq indicates a higher starting concentration of the transcript). Subsequently, we computed the abundance of each candidate gene relative to the internal control gene actin. These analyses were performed according to the instructions for relative quantification analyses in the Roche LightCycler® 96 manual to determine the ratio of target cDNA to actin (to account for potential differences in the starting concentration of cDNA among samples) based on raw Cq values and estimated cDNA concentrations. To test for changes in expression during dehydration and investigate sex differences in Dht gene expression, the relative ratio of target cDNA to actin (internal control gene) was analyzed in JMP®, Version 12 (SAS Institute Inc.). We used a mixed effects linear model to test the fixed effects of gene, hydration state, sex, and all second and third order interaction effects on the relative ratio of target:control expression. Contrasts were used to test for changes in expression during dehydration of each individual gene.

Data Availability
The unprocessed sequence data associated with this study are available in GenBank (accession numbers: SRR7348360 and SRR7348361). This Whole Genome Shotgun project (M_inflexa_v1.1) has been deposited at DDBJ/ENA/GenBank under the accession QLSQ00000000. The version described in this paper is version QLSQ01000000. The complete mitochondrial and chloroplast sequences are available at FigShare (https://doi.org/10.6084/m9.figshare.6639209.v1).

References
1. Kenrick, P & Crane, P.R. The origin and early evolution of plants on land. Nature 389, 33–39 (1997).
2. Bowman, J. L. et al. Insights into land plant evolution garnered from the Marchantia polymorpha genome. Cell 171, 287–304.e15 (2017).
3. Liggrone, R., Duckett, J. G. & Renzaglia, K. S. Major transitions in the evolution of early land plants: a bryological perspective. Annu. Bot. 109, 851–71 (2012).
4. Wellman, C. H., Osterloff, P. L. & Mohiuddin, U. Fragments of the earliest land plants. Nature 425, 282–285 (2003).
5. Morris, J. L. et al. The timescale of early land plant evolution. Proc. Natl. Acad. Sci. USA 115, E2274–E2283 (2018).
6. Leitch, I. J. & Leitch, A. R. Genome size diversity and evolution in land plants. in Plant Genome Diversity (eds Jones, R. N. & Langdon, T.) 307–322, https://doi.org/10.1007/978-3-7091-1160-4_2 (Springer, 2013).
7. Rensing, S. A. et al. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science (80-.). 319, 64–9 (2008).
8. Shaw, A. J. et al. The Sphagnum Genome Project. in 167–187, https://doi.org/10.1016/bs.abr.2016.01.003 (2016).
9. Szövényi, P. The Genome of the Model Species Anthoceros agrestis. Adv. Bot. Res. 78, 189–211 (2016).
10. Rensing, S. A. Why we need more non-seed plant models. New Phytol. 216, 355–360 (2017).
11. Oda, K. et al. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. Mol. Biol. Evol. 223, 1 (1999) and references therein.
12. Terasawa, K. et al. The Mitochondrial Genome of the Moss Physcomitrella patens Sheds New Light on Mitochondrial Evolution in Land Plants. Mol. Biol. Evol. 24, 699–709 (2007).
13. Wang, B., Xue, J., Li, L., Liu, Y. & Qiu, Y.-L. The complete mitochondrial genome sequence of the liverwort Pleurozia purpurea reveals extremely conservative mitochondrial genome evolution in liverworts. Curr. Genet. 55, 601–609 (2009).
14. Li, L., Wang, B., Liu, Y. & Qiu, Y.-L. The Complete Mitochondrial Genome Sequence of the Hornwort Megaceros aenigmaticus Shows a Mixed Mode of Conservative Yet Dynamic Evolution in Early Land Plant Mitochondrial Genomes. J. Mol. Evol. 68, 665–678 (2009).
15. Mysczyński, K. et al. The complete mitochondrial genome of the cryptic species C. of Aneura pinguis. Mitochondrial DNA Part A 28, 112–113 (2017).
16. Sugita, C., Kobayashi, Y., Aoki, S., Sugita, C. & Sugita, M. Complete chloroplast DNA sequence of the moss Physcomitrella patens: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. Nucleic Acids Res. 31, 3324–31 (2003).
61. Palmer, J. D. et al. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci. USA* **97**, 6960–6 (2000).
62. Palumbi, S. R. All males are not created equal: fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proc. Natl. Acad. Sci. USA* **96**, 12632–7 (1999).
63. Marks, R. A., Pike, B. & Nicholas Mcletchie, D. Genetic differences in water stress tolerance track environmental exposure and exhibit a fluctuating sexual dimorphism. *Rev. Ecoligia*.
64. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
65. Smith, J. J. et al. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* **50**, 270–277 (2018).
66. Hara, Y. et al. Optimizing and benchmarking de novo transcriptome sequencing: from library preparation to assembly evaluation. *BMC Genomics* **16**, 977 (2015).
67. Monte, I. et al. Ligand-receptor co-evolution shaped the jasmonate pathway in land plants. *Nat. Chem. Biol.* **14**, 480–488 (2018).
68. Okaizaki, N. et al. Novel factor highly conserved among eukaryotes controls sexual development in fission yeast. *Mol. Cell. Biol.* **18**, 887–95 (1998).
69. Wood, A. J. & Oliver, M. J. Translational control in plant stress: the formation of messenger ribonucleoprotein particles (mRNPs) in response to desiccation of Tortula ruralis gametophytes. *Plant J.* **18**, 359–370 (1999).
70. Scott, H. B. & Oliver, M. J. Accumulation and polysomal recruitment of transcripts in response to desiccation and rehydration of the moss Tortula ruralis. *J. Exp. Bot.* **45**, 577–583 (1994).
71. Schulz, P., Herde, M. & Romeis, T. Calcium-Dependent Protein Kinases: Hubs in Plant Stress Signaling and Development. *Plant Physiol.* **163** (2013).
72. Cheng, S.-H., Willmann, M. R., Chen, H.-C. & Sheen, J., Calcium Signaling through Protein Kinases. The Arabidopsis Calcium-Dependent Protein Kinase Gene Family. *Plant Physiol.* **129** (2002).
73. Doyle, J. J. & Doyle, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem.* **Bull.** **19**, 11–15 (1987).
74. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. (2010).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 214–214 (2014).
76. Rizk, G., Lavenier, D. & Chikhi, R. DSK: k-mer counting with very low memory usage. *Bioinformatics* **29**, 266–272 (2013).
77. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
78. Pruitt, K. D., Tatusova, T. & Maglott, R. D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D504–D508 (2005).
79. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–10 (1990).
80. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–86 (2007).
81. Sharma, N., Jung, C.-H., Bhalla, P. L. & Singh, M. B. RNA Sequencing Analysis of the Gametophyte Transcripom from the Liverwort, Marchantia polymorpha. *Plos One* **9**, e97497 (2014).
82. Tazazian, G. et al. Chromosomer: a reference-based genome arrangement tool for producing draft chromosome sequences. *Gigascience* **5**, 38 (2016).
83. Gnerre, S. et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **108**, 1513–1518 (2011).
84. Ohyama, K. et al. Structure and organization of Marchantia polymorpha chloroplast genome. I. Cloning and gene identification. *J. Mol. Biol.* **203**, 281–98 (1988).
85. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/arXiv:1303.3997 (2013).
86. Boisvert, S., Laviolette, F. & Corbeil, J. Ray: Simultaneous Assembly of Reads from a Mix of High-Throughput Sequencing Technologies. *J. Comput. Biol.* **17**, 1519–1533 (2010).
87. Assefa, S., Keane, T. M., Otto, T. D., Newbold, C. & Berriman, M. ARACAS: algorithm-based automatic conformation of assembled sequences. *Bioinformatics* **25**, 1968–1969 (2009).
88. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–402 (1997).
89. Walker, B. J. et al. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *Plos One* **9**, e112963 (2014).
90. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
91. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 88–96 (2008).
92. Berardini, T. Z. et al. The Arabidopsis information resource: Making and mining the ‘gold standard’ annotated reference plant genome. *Genesis* **53**, 474–85 (2015).
93. Harris, R. S. Improved pairwise alignment of genomic DNA. (Pennsylvania State University, 2007).
94. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
95. Nep, S. et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
96. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
97. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, 699–12 (2006).
98. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
99. Yang, Z. & Nielsen, R. Estimating Synonymous and Non synonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* **17**, 32–43 (2000).
100. Villanueva-Cañas, J. L., Laurie, S. & Albà, M. M. Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome Biol. Evol.* **5**, 457–67 (2013).
101. McCarthy, F. M. et al. AgBase: a functional genomics resource for agriculture. *BMC Genomics* **7**, 229 (2006).
102. Langmead, B., Trapnell, C., Pop, M. & Salberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.** **10**, R25 (2009).
103. Untergasser, A. et al. Primer3–new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).

**Acknowledgements**

Genome sequencing was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants Program (grant no. RGPIN-2014-05820) grant to R.C. Field work was funded by a University of Kentucky, Department of Biology’s Ribble grant to R.A.M. and the 2016 American Bryological and Lichenological Society’s Anderson and Crum Bryology Field Research Award to R.A.M. We thank the Flora Ribble Endowment,
University of Kentucky, Department of Biology for a fellowship to R.A.M. that enabled this work. We thank the University of Kentucky, College of Agriculture, Food and Environment for greenhouse space; the Wildlife Section, Forestry Division, Ministry of Agriculture, Land and Marine Resources of Trinidad and Tobago for collection and export permits; the Water and Sewage Authority for access to the research sites; and Andrea and Darryl McLetchie for logistical support in Trinidad. We thank John L. Bowman (Monash University) for helpful correspondence, and numerous colleagues who provided feedback on earlier versions of the manuscript.

Author Contributions
D.N.M. and Q.C. conceived of the study. D.N.M. and R.A.M. conducted DNA extraction and sex marker validation. C.J.G., J.J.S. and R.A.M. analyzed the data. D.N.M. and J.J.S., Q.C. and R.A.M. contributed to the interpretation of the data. R.A.M. wrote the manuscript. All authors reviewed and edited the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-45039-9.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019