bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title: Chromosome fusions shape an ancient UV sex chromosome system

Authors: Sarah B. Carey[1,§,‡], Jerry Jenkins[2], Adam C. Payton[1,3], Shenqiang Shu[4], John T. Lovell[2], Florian Maumus[5], Avinash Sreedasyam[2], George P. Tiley[6], Noe Fernandez-Pozo[7], Kerrie Barry[4], Cindy Chen[4], Mei Wang[4], Anna Lipzen[4], Chris Daum[4], Christopher A. Saski[8], Jordan C. McBreen[1], Roth E. Conrad[9], Leslie M. Kollar[1], Sanna Olsson[10], Sanna Huttunen[11], Jacob B. Landis[12], J. Gordon Burleigh[1], Norman J. Wickett[13], Matthew G. Johnson[14], Stefan A. Rensing[7], Jane Grimwood[2,4], Jeremy Schmutz[2,4], and Stuart F. McDaniel[1,*]

Affiliations:

[1]Department of Biology, University of Florida, Gainesville, FL, USA

[2]Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

[3]RAPiD Genomics, Gainesville, FL, USA

[4]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[5]Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France

[6]Department of Biology, Duke University, Durham, NC, USA

[7]Plant Cell Biology, University of Marburg, Marburg, Germany

[8]Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA

[9]School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[10]Department of Forest Ecology and Genetics, INIA-CIFOR, Madrid, Spain

[11]Department of Biology & Biodiversity Unit, University of Turku, Finland

[12]School of Integrative Plant Science, Section of Plant Biology and the L.H. Bailey Hortorium, Cornell University, Ithaca, NY, USA

[13]Negaunee Institute for Plant Conservation Science and Action, Chicago Botanic Garden, Glencoe, IL, USA

[14]Department of Biological Sciences, Texas Tech University, Lubbock, TX, USA

[§]Current address: Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL, USA

[‡]Current address: HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

[*]Corresponding author: stuartmcdaniel@ufl.edu

One Sentence Summary: Moss sex chromosomes retain thousands of broadly-expressed genes despite millions of years of suppressed recombination.

Abstract: Sex chromosomes occur in diverse organisms, but their structural complexity has often prevented evolutionary analyses. Here we use two chromosome-scale reference genomes of the moss *Ceratodon purpureus* to trace the evolution of the sex chromosomes in bryophytes. Comparative analyses show the moss genome comprises seven remarkably stable ancestral chromosomal elements. An exception is the sex chromosomes, which share thousands of broadly-expressed genes but lack any synteny. We show the sex chromosomes evolved over 300 million years ago and expanded via at least two distinct chromosomal fusions. These results link suppressed

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

recombination between the sex chromosomes with rapid structural change and the evolution of distinct transposable element compositions, and suggest haploid gene expression promotes the evolution of independent female and male gene-regulatory networks.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main Text:

Sex chromosomes arise when an ordinary pair of autosomes gains the capacity to determine sex (*1*). A defining characteristic of karyotypic sex-limited chromosomes (like the mammalian Y) is suppressed recombination, which links alleles that promote one sexual function. It is widely believed the lack of recombination predisposes such chromosomes to degeneration and gene-loss because natural selection is less effective in regions of low recombination (*2*, *3*). The processes by which sex chromosomes gain new genes, however, are less well understood. Multiple neutral or adaptive processes can lead to fusions between sex chromosomes and autosomes, forming neo-sex chromosomes (*4*, *5*). Such fusions are well known in animals (*6*–*10*), but most animal sex chromosomes rapidly degenerate to the point where few homologous genes remain (*2*), providing little data to evaluate the forces shaping these events.

Plant sex chromosomes show promise as models for the evolution of sex chromosome fusions because gene loss is apparently slowed by strong purifying selection at the haploid stage (e.g., pollen) (*11*), meaning more genes remain for comparative analyses. However, flowering plant sex chromosomes in even closely-related genera may share no genes in common, making such analyses difficult (*12*). In contrast, bryophyte sex chromosome systems may be more broadly shared (*13*, *14*), although the origins and homology of these systems are unknown. Unlike the diploid XY and ZW systems, where recombination is suppressed only on the Y or W, bryophytes typically have a UV system in which haploid females possess a U chromosome while haploid males possess a V (*15*). The U and V pair at meiosis in the diploid sporophyte, like their X-Y counterparts, and segregate to female and male spores, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

However, because there is no homogametic sex, recombination is suppressed on both sex chromosomes (i.e., all diploids are UV, meaning UU and VV individuals are never formed), and the UV system lacks the hemizygosity found in diploid sex chromosomes. Therefore, the asymmetric gene loss that characterizes the X-Y pair is unlikely between the U and V (*16*, *17*). However, processes driving sex chromosome fusions in XY (or ZW) systems, including degeneration (due to suppressed recombination or sex-limited inheritance) or various forms of antagonistic selection (e.g. sex-ratio distortion, (*18*); sexual dimorphism (*19*); or parent-offspring conflict, (*20*)), may also act on bryophyte sex chromosomes.

The moss ancestral chromosomal elements.

To reconstruct the history of the bryophyte sex chromosomes, we first assembled chromosome-scale genomes of a male (R40) and a female (GG1) isolate of the moss *Ceratodon purpureus*. Using a combination of Illumina, Bacterial Artificial Chromosomes (BACs), PacBio, and Dovetail Hi-C (*21*) (Table S1-8), the version 1.0 genome assembly of R40 comprises 358 Megabases (Mb) in 601 contigs (N50 1.4 Mb), with 98.3% of the assembled sequence in the largest 13 pseudomolecules, corresponding to the 13 chromosomes in the *C. purpureus* karyotype (*22*) (Fig.1A). The version 1.0 GG1 assembly is 349.5 Mb in 558 contigs (N50 1.4 Mb), with 97.9% of assembled sequence in the largest 13 pseudomolecules. The two genome lines, GG1 and R40, were isolated from distant localities (Gross Gerungs, Austria and Rensselaer, New York, USA, respectively) (*23*), and are only partially interfertile (*24*), potentially as a result of the numerous structural differences between the genomes (Fig. 1B). Using >1.5 billion

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RNA-seq reads for each of the genome lines and additional *de novo* assemblies of other *C. purpureus* isolates (*21*) (Table S9), we annotated 31,482 genes on the R40 assembly and 30,425 on GG1 (BUSCO v3.0 of 69% using Embryophyte; 96.7 and 96.4%, respectively using Eukaryote; comparable to those for *Physcomitrium patens*) (*25*).

We performed a self-synteny analysis and found clear homeologous chromosome pairs resulting from an ancient whole-genome duplication (WGD) event (*21*) (Fig. 1C, S3), consistent with previous transcriptomic analyses (*23*, *26*) and our own Ks-based analyses (*21*) (Fig. S4; Table S10-11). Remarkably, we also identified abundant synteny between *C. purpureus* and *P. patens* (*27*) (Fig. 1D), which diverged over 200 million years ago (*28*). The inferred karyotype from the ancestor elements, and therefore the ancestor of most extant mosses, consisted of seven chromosomes (*27*), which we refer to as Ancestral Elements A-G (Fig. 1D). These data suggest major parts of the gene content of moss ancestral chromosome elements are stable over hundreds of millions of years, similar to the “Muller Elements” in *Drosophila* flies (*29*).

The *C. purpureus* sex chromosomes are large and gene rich.

The major exception to the long-term genomic stability observed in *C. purpureus* is the sex chromosomes, which possess no obviously syntenic regions between the U and V or with the autosomes (Fig. 1B, C). Notably, the sex chromosomes are by far the largest chromosomes in the genome assembly, containing ~30% of each genome (110.5 Mb on the V in R40, and 112.2 Mb on the U in GG1; Fig. 1-2) and are each ~3.8 times bigger than the largest autosome (e.g., R40 chromosome one is 29 Mb). One possible reason

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for their lack of synteny and large size is an increase in transposable elements (TEs), which often accumulate in regions of low recombination and may promote structural rearrangements. The 12 autosomal chromosomes in *C. purpureus* contain on average 46.4% TEs (*21*) (ranging between 43.2 and 49.9%; Fig. 2; Table S12), whereas the U and V sex chromosomes contain 78.2% and 81.9%, respectively, similar to the non-recombining Y or W sex chromosomes in other systems (*30*). While some TEs have fairly-even distributions across all chromosomes (e.g., Copia and TIR), strikingly, the U and V chromosomes are enriched for very different classes of repeats compared to each other and the autosomes. The U is enriched for Sola, hAT and Gypsy, and the V is enriched in a previously undescribed superfamily of cut-and-paste DNA transposons, which we refer to as Lanisha elements (Fig. 2, S5-6; Table S13-14). The distribution of repeats in *C. purpureus* and the Hi-C contact map (Fig. 1A) together suggest the chromatin structure in the nucleus that permits TEs to be shared among the autosomes is absent between the autosomes and the sex chromosomes (*31*), highlighting the enigmatic isolation of the sex chromosomes in the nucleus.

Both the U and the V in *C. purpureus* contain many genes (3,450 and 3,411, respectively), which stands in stark contrast to the non-recombining mammalian Y chromosome, or even other UV systems like the liverwort *Marchantia polymorpha,* the green alga *Volvox carteri*, or the brown alga *Ectocarpus silicius*, which typically contain an order of magnitude fewer (*14*, *32*, *33*). Despite the fact that the U and V sex chromosomes have lower gene density than the autosomes (Fig. 2), they nonetheless represent ~12% of the gene content in this species for both sexes. This observation

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

supports theory suggesting suppressed recombination alone is insufficient to drive gene loss and degeneration on UV sex chromosomes (*17*).

Because sex-linked genes may be involved in sex-specific development, we expected the sex chromosomes to contain transcription factors (TF) or trancriptional regulators (TR) that control these processes. Indeed, some TF/TRs are only found on the UV sex chromosomes (*21*) (e.g., HD DDT, Med7, and SOH1; Table S15). However, the overall representation of most TF/TR families on the sex chromosomes is low (6% of identified TF/TRs found on the U and 5% on V, Fig. 2; Table S15) compared to the proportion of genome sequence (30%) or genes (12%) residing on the sex chromosomes. Indeed, the sex chromosomes do contain potential key regulators of sexual differentiation, such as a U-linked homolog in the RWP-RK gene family, which in green algae is a mating-type factor and in *M. polymorpha* regulates egg cell formation and maintenance (*32*, *34*). However, transcriptomic data from juvenile and mature-stage tissues in six additional isolates (*21*) (Table S9,16), show more than 1,700 U and V-linked genes were expressed at a mean count ≥1, including components of the cytoskeleton (e.g., Tubulin) and DNA repair complexes (e.g., RAD51). Thus, numerous structural genes, rather than just regulators, may be retained on UV sex chromosomes. Further, the genes on the U and V sex chromosomes were enriched for different GO and KEGG terms (Fig, S7; Table S17-20) and their expression was largely correlated with other sex-linked genes (Fig S8; Table S21-22).

Moss sex chromosomes are ancient but evolutionarily dynamic.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The abundance of genes on the *C. purpureus* sex chromosomes provide an unusually large dataset for molecular evolutionary reconstruction. Critically, the times to the most-recent common ancestor between homologous genes on the U and V chromosomes allows us to estimate when these genes became sex-linked. To accomplish this, we used genome annotations or *de novo* assembled genes from RNA-seq for 41 mosses, 21 liverworts, one lycophyte, and two ferns (Table S23), including sex-specific RNA-seq for the moss *Brachythecium rivulare*, to infer when sex-linkage occured for genes annotated to the *C. purpureus* U and V chromosomes.

We clustered all annotated *C. purpureus* peptides, including the 3,450 U and 3,411 V-linked, with the other species using Orthofinder (*35*) and found 2,450 homologous U-V gene pairs. Using a phylogenetic approach with stringent inclusion criteria (*21*), we built 744 gene trees, 402 with U and V-linked homologs (the remaining were U or V-specific; Table S24). If a gene became sex-linked recently, we expect to find the *C. purpureus* U and V copies sister to each other, while even closely-related species possess autosomal copies (*24*). However, in an older capture or recombination suppression event, we expect to find the U-linked copies among different species to share a common ancestor with one another before they share an ancestor with the V-linked homologs.

We found most genes became sex-linked in the *C. purpureus* lineage, after the divergence from *Syntrichia princeps* (the closest relative for which we have sequence data; mean silent-site divergence (Ks) = 0.16; Fig. 3A; Table S25). However, 13 genes diverged at the base of the Dicraniidae (mean Ks = 0.85; Fig. 3A; Table S25), and three genes diverged prior to the split between the two diverse clades Bryidae and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dicraniidae (mean Ks = 1.64; Fig. 3A; Table S25). The most ancient U-V coalescence (a Zinc finger Ran binding protein of unknown function) was prior to the split between *Buxbaumia aphylla* and the remaining Bryopsida, ~300 MYA (based on previous fossil-calibrated, relaxed-clock analyses from (*28*); Ks = 2.8; Fig. 3A, S9; Table S25). The ancient divergence between the U-V pair unambiguously shows dioecy arose far earlier than previously inferred based on ancestral-state reconstructions of the sexual system in mosses (*13*). The actual sex-determining gene may be older than this, but because our analysis depends on identifying both U and V-linked orthologs, we could not evaluate the age of sex-specific genes (e.g., the ampliconic genes on the mammalian Y (*36*)).

A classic signature of gene-capture on sex chromosomes is the presence of strata, where neighboring genes added in the same recombination suppression event have a similar Ks between sex-linked copies. Finding many contiguous genes with a similar Ks is consistent with a large structural change, as seen in the strata on the human X-Y pair (*37*). In contrast, a more gradual change in U-V divergence has been proposed to reflect many smaller gene-capture events (*24*). To test for evidence of strata on the UV sex chromosomes we searched for patterns of U–V Ks across the chromosomes (*21*). In contrast to either the stepwise or gradual expansion scenarios outlined above, we found Ks was not associated with gene order on the *C. purpureus* sex chromosomes (Fig. 3B; Table S25). Even genes with very low Ks values, presumably from the most-recent capture events, were found across the entirety of the U or V chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

One possible mechanism to explain the lack of coincidence between physical position and Ks is recurrent gene conversion during the diploid UV phase that homogenizes the U and V-linked copies of homologous genes. Consistent with this phenomenon, we found a case in the Bryidae where the female copy of one sex-linked gene was converted to the male copy but subsequently diverged again (Fig. S10). Nevertheless, even frequent gene conversion cannot explain the lack of conserved gene order between the U and V chromosomes. The lack of U-V synteny, however, does not rule out a history of large fusions onto the moss sex chromosomes if there were many subsequent rearrangements (Fig. 3B).

Cryptic fusions on the moss UV sex chromosomes.

Here we employ a phylogenomic approach to search for cryptic evolutionary strata between the rearranged U and V sex chromosomes (*21*). In the self-synteny analysis, chromosomes 5 (Ancestral Element D) and 9 (Ancestral Element B; Fig. 1C, 1D) conspicuously lacked a homeologous chromosome. If the missing chromosomes were lost due to aneuploidy, we would expect to find gene trees containing only a single *C. purpureus* copy of genes from chromosomes 5 or 9, along with the respective *P. patens* chromosomes from the same Ancestral Element (Fig. 1D). In contrast, gene trees that additionally include a U or V-linked copy provide evidence the missing homeolog to chromosome 5 or 9 was fused to the sex chromosome (Fig. 3C). Indeed, when we examined trees for the recently-captured UV homologs (i.e, those only in *C. purpureus*), we found 309 gene trees with Ancestral Element D paralogs in *P. patens* (Table S26, including 47 paralogous genes from *C. purpureus* chromosome 5, Table S27). The most

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

parsimonious explanation for this observation is one of the Ancestral Element D duplicate chromosomes (Fig. 1D) fused to the sex chromosome, but extensive rearrangements made it impossible to detect this fusion using synteny-based approaches (Fig. 3B).

To further evaluate the cryptic neo-sex chromosome hypothesis, we surveyed the gene trees from the other phylogenomically defined UV expansions (Fig. 3A). All but one of the genes from the capture event in the ancestor of the Dicranidae mapped to Element B, and presumably correspond to the missing chromosome 9 homeolog (Fig. 1D; Table S26). Additionally, the two genes with clear *P. patens* orthologs from the oldest two capture events correspond to Element A (Table S26). Together these phylogenomic data highlight the role of rare chromosome fusions in the evolution of sex chromosome gene content and also point to Element A as the ancestral sex chromosome in mosses.

Liverworts show a similar evolutionary history as mosses.

To test whether gene acquisition and lack of U-V synteny are shared across Setaphyte (i.e., moss and liverwort) sex chromosomes, we applied our phylogenomic method to detect gene capture on the sex chromosomes of *M. polymorpha* (*14*, *21*). Using orthologous U-V genes, we found evidence of four capture events on the *M. polymorpha* sex chromosomes, with the oldest diverging ~400 MYA, prior to the split of Marchantiidae and Pelliidae (date following (*28*); Fig. 3, S11). As in *C. purpureus*, we found no evidence of syntenic strata when we compare Ks between the U and V-linked orthologs (based on the linear order in which they are annotated on the chromosomes;

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S28). However, using synteny-agnostic gene trees we identified the *P. patens* homologs of 18 *M. polymorpha* sex-linked genes. Ten of these genes, across all four capture events, shared a common ancestor on the moss Ancestral Element A, suggesting this element played a key role in sex determination early in the history of the liverwort lineage as well (Table S29). Notably, six of the remaining genes across capture events have homologous copies on Ancestral Elements B and D, similar to mosses (Table S29). These observations lead us to the remarkable conclusion the same ancestral elements were recruited to the sex chromosome after the divergence between mosses and liverworts and, moreover, hint the Setaphytes may share a common sex-determining system which originated before the split between these lineages, ~500 MYA.

We do find clear evidence of more recent convergent recruitment of numerous genes to the non-recombining U and V in *C. purpureus* and *M. polymorpha*. For example, the V-specific ABC1 family CepurR40.VG273600 is homologous to MapolyY_A0026 and MapolyY_A0027 (Fig. S12). With our data, we cannot determine when this gene was captured in *C. purpureus* using gene trees, because the U-linked copy was lost in both species. However, because it has a chromosome 5 homolog (CepurR40.5G184800) it was likely present on Element D and therefore captured recently. Additionally, a different gene (Nudix hydrolase related) based on the topology was clearly captured independently in *C. purpureus* (CepurR40.VG335300; CepurGG1.UG335300) and *M. polymopha* (Mapoly0018s0016; MapolyY_B0027; Fig. S13). Further, the autosomal cis-acting sexual dimorphism switch *MpFGMYB* (*38*) that promotes female development in *M. polymorpha* is also orthologous to recently-

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

captured U (CepurGG1.UG291100) and V-linked (CepurR40.VG015400) copies in *C. purpureus* (Fig. S14). The parallel evolution of sex-linkage of genes with clear roles in sexual differentiation in both mosses and liverworts suggests sexual antagonism plays an important role in either the evolution of bryophyte sex chromosome translocations (*4*, *5*) or the subsequent evolution of translocated genes.

Overall, the gene-rich U and V chromosomes of *C. purpureus* strongly suggest haploid gene expression slows sex chromosome degeneration, even over millions of years of suppressed recombination. However, the lack of meiotic recombination is associated with massive structural variation between the U and V, and highly differentiated TE accumulation, which stands in stark contrast to the conserved synteny and homogeneous TE composition of the autosomes. Moreover, the remarkable antiquity of dioecy in Setophytes more closely mirrors the evolution of sexual systems in animals rather than flowering plants, where hermaphroditism is the norm (*39*, *40*). The phylogenetic distribution of sexual system transitions, combined with the abundance of sex-linked genes and evidence for sex-specific gene regulatory networks, position bryophytes as a powerful model for studying factors shaping the evolution and maintenance of dioecy.

References and Notes:

1. J. J. Bull, *Evolution of sex determining mechanisms* (The Benjamin/Cummings Publishing Company, Inc., 1983).

2. B. Charlesworth, D. Charlesworth, The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000).

3. D. Bachtrog, Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

4. G. S. van Doorn, M. Kirkpatrick, Turnover of sex chromosomes induced by sexual conflict. *Nature*. **449**, 909–912 (2007).

5. O. Blaser, C. Grossen, S. Neuenschwander, N. Perrin, Sex-chromosome turnovers induced by deleterious mutation load. *Evolution*. **67**, 635–645 (2013).

6. C. L. R. Delgado, P. D. Waters, C. Gilbert, T. J. Robinson, J. A. M. Graves, Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Res.* **17**, 917–926 (2009).

7. B. Vicoso, D. Bachtrog, Numerous transitions of sex chromosomes in Diptera. *PLoS Biol.* **13**, e1002078 (2015).

8. R. P. Meisel, P. U. Olafson, F. D. Guerrero, K. Konganti, J. B. Benoit, High rate of sex chromosome turnover in muscid flies. *bioRxiv* (2019), p. 655845.

9. T. Myosho, Y. Takehana, S. Hamaguchi, M. Sakaizumi, Turnover of Sex Chromosomes in Celebensis Group Medaka Fishes. *G3* . **5**, 2685–2691 (2015).

10. W. J. Gammerdinger, T. D. Kocher, Unusual Diversity of Sex Chromosomes in African Cichlid Fishes. *Genes* . **9** (2018), doi:10.3390/genes9100480.

11. M. V. Chibalina, D. A. Filatov, Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr. Biol.* **21**, 1475–1479 (2011).

12. D. Charlesworth, Plant Sex Chromosomes. *Annu. Rev. Plant Biol.* **67**, 397–420 (2016).

13. S. F. McDaniel, J. Atwood, J. G. Burleigh, Recurrent evolution of dioecy in bryophytes. *Evolution*. **67**, 567–572 (2013).

14. J. L. Bowman, T. Kohchi, K. T. Yamato, J. Jenkins, S. Shu, K. Ishizaki, S. Yamaoka, R. Nishihama, Y. Nakamura, F. Berger, C. Adam, S. S. Aki, F. Althoff, T. Araki, M. A. Arteaga-Vazquez, S. Balasubrmanian, K. Barry, D. Bauer, C. R. Boehm, L. Briginshaw, J. Caballero-Perez, B. Catarino, F. Chen, S. Chiyoda, M. Chovatia, K. M. Davies, M. Delmans, T. Demura, T. Dierschke, L. Dolan, A. E. Dorantes-Acosta, D. M. Eklund, S. N. Florent, E. Flores-Sandoval, A. Fujiyama, H. Fukuzawa, B. Galik, D. Grimanelli, J. Grimwood, U. Grossniklaus, T. Hamada, J. Haseloff, A. J. Hetherington, A. Higo, Y. Hirakawa, H. N. Hundley, Y. Ikeda, K. Inoue, S.-I. Inoue, S. Ishida, Q. Jia, M. Kakita, T. Kanazawa, Y. Kawai, T. Kawashima, M. Kennedy, K. Kinose, T. Kinoshita, Y. Kohara, E. Koide, K. Komatsu, S. Kopischke, M. Kubo, J. Kyozuka, U. Lagercrantz, S.-S. Lin, E. Lindquist, A. M. Lipzen, C.-W. Lu, E. De Luna, R. A. Martienssen, N. Minamino, M. Mizutani, M. Mizutani, N. Mochizuki, I. Monte, R. Mosher, H. Nagasaki, H. Nakagami, S. Naramoto, K. Nishitani, M. Ohtani, T. Okamoto, M. Okumura, J. Phillips, B. Pollak, A. Reinders, M. Rövekamp, R. Sano, S. Sawa, M. W. Schmid, M. Shirakawa, R. Solano, A. Spunde, N. Suetsugu, S. Sugano, A. Sugiyama, R. Sun, Y. Suzuki, M. Takenaka, D. Takezawa, H. Tomogane, M. Tsuzuki, T. Ueda, M. Umeda, J. M. Ward, Y. Watanabe, K. Yazaki, R. Yokoyama, Y. Yoshitake, I. Yotsui, S. Zachgo, J. Schmutz, Insights into Land Plant Evolution Garnered from the Marchantia polymorpha Genome. *Cell*. **171**, 287–304.e15 (2017).

15. D. Bachtrog, M. Kirkpatrick, J. E. Mank, S. F. McDaniel, J. C. Pires, W. Rice, N.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Valenzuela, Are all sex chromosomes created equal? *Trends Genet.* **27**, 350–357 (2011).

16. J. J. Bull, Sex Chromosomes in Haploid Dioecy: A Unique Contrast to Muller's Theory for Diploid Dioecy. *Am. Nat.* **112**, 245–250 (1978).

17. S. Immler, S. P. Otto, The evolution of sex chromosomes in organisms with separate haploid sexes. *Evolution*. **69**, 694–708 (2015).

18. T. E. Norrell, K. S. Jones, A. C. Payton, S. F. McDaniel, Meiotic sex ratio variation in natural populations of Ceratodon purpureus (Ditrichaceae). *Am. J. Bot.* **101**, 1572–1576 (2014).

19. S. F. McDaniel, Genetic correlations do not constrain the evolution of sexual dimorphism in the moss Ceratodon purpureus. *Evolution*. **59**, 2353–2361 (2005).

20. D. Haig, A. Wilczek, Sexual conflict and the alternation of haploid and diploid generations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **361**, 335–343 (2006).

21. Materials and methods are available as supporting material on *Science* Online.

22. R. Fritsch, Index to bryophyte chromosome counts (1991) (available at https://www.schweizerbart.de/publications/detail/isbn/9783443620127/%23).

23. P. Szövényi, P.-F. Perroud, A. Symeonidi, S. Stevenson, R. S. Quatrano, S. A. Rensing, A. C. Cuming, S. F. McDaniel, De novo assembly and comparative analysis of the Ceratodon purpureus transcriptome. *Mol. Ecol. Resour.* **15**, 203–215 (2015).

24. S. F. McDaniel, K. M. Neubig, A. C. Payton, R. S. Quatrano, D. J. Cove, RECENT GENE-CAPTURE ON THE UV SEX CHROMOSOMES OF THE MOSS CERATODON PURPUREUS : SEX CHROMOSOME EVOLUTION IN C. PURPUREUS. *Evolution*. **110** (2013), doi:10.1111/evo.12165.

25. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. **31**, 3210–3212 (2015).

26. One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. **574**, 679–685 (2019).

27. D. Lang, K. K. Ullrich, F. Murat, J. Fuchs, J. Jenkins, F. B. Haas, M. Piednoel, H. Gundlach, M. Van Bel, R. Meyberg, Others, The Physcomitrella patens chromosome-scale assembly reveals moss genome structure and evolution. *Plant J.* **93**, 515–533 (2018).

28. B. Laenen, B. Shaw, H. Schneider, B. Goffinet, E. Paradis, A. Désamoré, J. Heinrichs, J. C. Villarreal, S. R. Gradstein, S. F. McDaniel, D. G. Long, L. L. Forrest, M. L. Hollingsworth, B. Crandall-Stotler, E. C. Davis, J. Engel, M. Von Konrat, E. D. Cooper, J. Patiño, C. J. Cox, A. Vanderpoorten, A. J. Shaw, Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nat. Commun.* **5**, 5134 (2014).

29. S. W. Schaeffer, Muller "Elements" in Drosophila: How the Search for the Genetic Basis for Speciation Led to the Birth of Comparative Genomics. *Genetics*. **210**, 3–13 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

30. R. Bergero, D. Charlesworth, The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* **24**, 94–102 (2009).

31. S. A. Montgomery, Y. Tanizawa, B. Galik, N. Wang, T. Ito, T. Mochizuki, S. Akimcheva, J. L. Bowman, V. Cognat, L. Maréchal-Drouard, H. Ekker, S.-F. Hong, T. Kohchi, S.-S. Lin, L.-Y. D. Liu, Y. Nakamura, L. R. Valeeva, E. V. Shakirov, D. E. Shippen, W.-L. Wei, M. Yagura, S. Yamaoka, K. T. Yamato, C. Liu, F. Berger, Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Curr. Biol.* **30**, 573–588.e7 (2020).

32. P. Ferris, B. J. S. C. Olson, P. L. De Hoff, S. Douglass, D. Casero, S. Prochnik, S. Geng, R. Rai, J. Grimwood, J. Schmutz, I. Nishii, T. Hamaji, H. Nozaki, M. Pellegrini, J. G. Umen, Evolution of an expanded sex-determining locus in Volvox. *Science*. **328**, 351–354 (2010).

33. S. Ahmed, J. M. Cock, E. Pessia, R. Luthringer, A. Cormier, M. Robuchon, L. Sterck, A. F. Peters, S. M. Dittami, E. Corre, M. Valero, J.-M. Aury, D. Roze, Y. Van de Peer, J. Bothwell, G. A. B. Marais, S. M. Coelho, A haploid system of sex determination in the brown alga Ectocarpus sp. *Curr. Biol.* **24**, 1945–1957 (2014).

34. M. Rövekamp, J. L. Bowman, U. Grossniklaus, Marchantia MpRKD Regulates the Gametophyte-Sporophyte Transition by Keeping Egg Cells Quiescent in the Absence of Fertilization. *Curr. Biol.* **26**, 1782–1789 (2016).

35. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

36. H. Skaletsky, T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier, L. G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.-F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.-P. Yang, R. H. Waterston, R. K. Wilson, S. Rozen, D. C. Page, The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. **423**, 825–837 (2003).

37. B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science*. **286**, 964–967 (1999).

38. T. Hisanaga, K. Okahashi, S. Yamaoka, T. Kajiwara, R. Nishihama, M. Shimamura, K. T. Yamato, J. L. Bowman, T. Kohchi, K. Nakajima, A cis-acting bidirectional transcription switch controls sexual dimorphism in the liverwort. *EMBO J.* **38** (2019) (available at https://www.embopress.org/doi/abs/10.15252/embj.2018100240).

39. S. M. Eppley, L. K. Jesson, Moving to mate: the evolution of separate and combined sexes in multicellular organisms. *J. Evol. Biol.* **21**, 727–736 (2008).

40. D. A. Sasson, J. F. Ryan, A reconstruction of sexual modes throughout animal evolution. *BMC Evol. Biol.* **17**, 242 (2017).

41. S. F. McDaniel, A. J. Shaw, Selective sweeps and intercontinental migration in the cosmopolitan moss Ceratodon purpureus (Hedw.) Brid. *Mol. Ecol.* **14**, 1121–1132

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(2005).

42. M. Luo, R. A. Wing, An improved method for plant BAC library construction. *Methods Mol. Biol.* **236**, 3–20 (2003).

43. B. Ewing, L. Hillier, M. C. Wendl, P. Green, Base-Calling of Automated Sequencer Traces UsingPhred. I. Accuracy Assessment. *Genome Res.* **8**, 175–185 (1998).

44. B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).

45. D. Gordon, C. Abajian, P. Green, Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).

46. P.-F. Perroud, F. B. Haas, M. Hiss, K. K. Ullrich, A. Alboresi, M. Amirebrahimi, K. Barry, R. Bassi, S. Bonhomme, H. Chen, Others, The Physcomitrella patens gene atlas project: large-scale RNA-seq based expression data. *Plant J.* **95**, 168–182 (2018).

47. D. J. Cove, P.-F. Perroud, A. J. Charron, S. F. McDaniel, A. Khandelwal, R. S. Quatrano, Culturing the moss Physcomitrella patens. *Cold Spring Harb. Protoc.* **2009**, db.prot5136 (2009).

48. E. H. Leder, J. M. Cano, T. Leinonen, R. B. O'Hara, M. Nikinmaa, C. R. Primmer, J. Merilä, Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. *Mol. Biol. Evol.* **27**, 1495–1503 (2010).

49. C.-L. Xiao, Y. Chen, S.-Q. Xie, K.-N. Chen, Y. Wang, Y. Han, F. Luo, Z. Xie, MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods*. **14**, 1072–1074 (2017).

50. C.-S. Chin, D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E. E. Eichler, S. W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods*. **10**, 563–569 (2013).

51. N. C. Durand, M. S. Shamim, I. Machol, S. S. P. Rao, M. H. Huntley, E. S. Lander, E. L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. **3**, 95–98 (2016).

52. W. J. Kent, BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664 (2002).

53. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at http://arxiv.org/abs/1303.3997).

54. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

55. T. D. Wu, J. Reeder, M. Lawrence, G. Becker, M. J. Brauer, GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. *Methods Mol. Biol.* **1418**, 283–334 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

56. J. Krumsiek, R. Arnold, T. Rattei, Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics*. **23**, 1026–1028 (2007).

57. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr, L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).

58. A. F. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013--2015 (2015).

59. A. F. A. Smit, R. Hubley, P. Green, RepeatModeler Open-1.0. 2008--2015. *Seattle, USA: Institute for Systems Biology. Available from: httpwww. repeatmasker. org, Last Accessed May*. **1**, 2018 (2015).

60. A. A. Salamov, V. V. Solovyev, Ab initio gene finding in Drosophila genomic DNA. *Genome Res.* **10**, 516–522 (2000).

61. G. S. C. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*. **6**, 31 (2005).

62. K. J. Hoff, S. Lange, A. Lomsadze, M. Borodovsky, M. Stanke, BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. **32**, 767–769 (2016).

63. A. Zwaenepoel, Y. Van de Peer, wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics*. **35**, 2153–2155 (2019).

64. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).

65. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).

66. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

67. N. Goldman, Z. Yang, A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).

68. S. Proost, J. Fostier, D. De Witte, B. Dhoedt, P. Demeester, Y. Van de Peer, K. Vandepoele, i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11–e11 (2012).

69. G. P. Tiley, M. S. Barker, J. G. Burleigh, Assessing the Performance of Ks Plots for Detecting Ancient Whole Genome Duplications. *Genome Biol. Evol.* **10**, 2882–2898 (2018).

70. S. A. Rensing, J. Ick, J. A. Fawcett, D. Lang, A. Zimmer, Y. Van de Peer, R. Reski, An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. *BMC Evol. Biol.* **7**, 130 (2007).

71. J. T. Lovell, J. Jenkins, D. B. Lowry, S. Mamidi, A. Sreedasyam, X. Weng, K. Barry, J. Bonnette, B. Campitelli, C. Daum, S. P. Gordon, B. A. Gould, A. Khasanova, A.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Lipzen, A. MacQueen, J. D. Palacio-Mejía, C. Plott, E. V. Shakirov, S. Shu, Y. Yoshinaga, M. Zane, D. Kudrna, J. D. Talag, D. Rokhsar, J. Grimwood, J. Schmutz, T. E. Juenger, The genomic landscape of molecular responses to natural drought stress in Panicum hallii. *Nat. Commun.* **9**, 5213 (2018).

72. T. Flutre, E. Duprat, C. Feuillet, H. Quesneville, Considering transposable element diversification in de novo annotation approaches. *PLoS One*. **6**, e16526 (2011).

73. H. Quesneville, C. M. Bergman, O. Andrieu, D. Autard, D. Nouaud, M. Ashburner, D. Anxolabehere, Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* **1** (2005) (available at https://www.ncbi.nlm.nih.gov/pmc/articles/pmc1185648/).

74. G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).

75. C. Hoede, S. Arnoux, M. Moisset, T. Chaumier, O. Inizan, V. Jamilloux, H. Quesneville, PASTEC: an automatic transposable element classification tool. *PLoS One*. **9**, e91929 (2014).

76. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*. **20**, 473 (2019).

77. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).

78. Y.-W. Yuan, S. R. Wessler, The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7884–7889 (2011).

79. H. Tsubouchi, G. S. Roeder, The Mnd1 protein forms a complex with hop2 to promote homologous chromosome pairing and meiotic double-strand break repair. *Mol. Cell. Biol.* **22**, 3078–3088 (2002).

80. D. Lang, B. Weiche, G. Timmerhaus, S. Richardt, D. M. Riaño-Pachón, L. G. G. Corrêa, R. Reski, B. Mueller-Roeber, S. A. Rensing, Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol. Evol.* **2**, 488–503 (2010).

81. P. K. I. Wilhelmsson, C. Mühlich, K. K. Ullrich, S. A. Rensing, Comprehensive Genome-Wide Classification Reveals That Many Plant-Specific Transcription Factors Evolved in Streptophyte Algae. *Genome Biol. Evol.* **9**, 3384–3397 (2017).

82. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

83. R. C. Team, R: a language and environment for statistical computing (version 3.5. 3)[software] (2019).

84. B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*. **33**, 3088–3090 (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

85. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

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

87. K. C. Olney, S. M. Brotman, V. Valverde-Vesling, J. Andrews, M. A. Wilson, Aligning RNA-Seq reads to a sex chromosome complement informed reference genome increases ability to detect sex differences in gene expression. *BioRxiv*, 668376 (2019).

88. M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell, S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

89. M. Love, S. Anders, W. Huber, Differential analysis of count data--the DESeq2 package. *Genome Biol.* **15**, 10–1186 (2014).

90. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008).

91. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

92. G. Csardi, T. Nepusz, Others, The igraph software package for complex network research. *InterJournal, complex systems*. **1695**, 1–9 (2006).

93. F. Briatte, ggnetwork: Geometries to Plot Networks with'ggplot2'. *R package version 0. 5*. **1** (2016).

94. A. Alexa, J. Rahnenfuhrer, topGO: enrichment analysis for gene ontology. *R package version*. **2**, 2010 (2010).

95. G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, S. Wang, GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. **26**, 976–978 (2010).

96. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

97. J. A. Banks, T. Nishiyama, M. Hasebe, J. L. Bowman, M. Gribskov, C. dePamphilis, V. A. Albert, N. Aono, T. Aoyama, B. A. Ambrose, N. W. Ashton, M. J. Axtell, E. Barker, M. S. Barker, J. L. Bennetzen, N. D. Bonawitz, C. Chapple, C. Cheng, L. G. G. Correa, M. Dacre, J. DeBarry, I. Dreyer, M. Elias, E. M. Engstrom, M. Estelle, L. Feng, C. Finet, S. K. Floyd, W. B. Frommer, T. Fujita, L. Gramzow, M. Gutensohn, J. Harholt, M. Hattori, A. Heyl, T. Hirai, Y. Hiwatashi, M. Ishikawa, M. Iwata, K. G. Karol, B. Koehler, U. Kolukisaoglu, M. Kubo, T. Kurata, S. Lalonde, K. Li, Y. Li, A. Litt, E. Lyons, G. Manning, T. Maruyama, T. P. Michael, K. Mikami, S. Miyazaki, S.-I. Morinaga, T. Murata, B. Mueller-Roeber, D. R. Nelson, M. Obara, Y. Oguri, R. G. Olmstead, N. Onodera, B. L. Petersen, B. Pils, M. Prigge, S. A. Rensing, D. M. Riaño-Pachón, A. W. Roberts, Y. Sato, H. V. Scheller, B. Schulz, C. Schulz, E. V. Shakirov, N. Shibagaki, N. Shinohara, D. E. Shippen, I. Sørensen, R. Sotooka, N. Sugimoto, M. Sugita, N. Sumikawa, M. Tanurdzic, G. Theissen, P. Ulvskov, S. Wakazuki, J.-K. Weng, W. W. G. T. Willats, D. Wipf, P. G. Wolf, L. Yang, A.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

D. Zimmer, Q. Zhu, T. Mitros, U. Hellsten, D. Loqué, R. Otillar, A. Salamov, J. Schmutz, H. Shapiro, E. Lindquist, S. Lucas, D. Rokhsar, I. V. Grigoriev, The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science*. **332**, 960–963 (2011).

98. F.-W. Li, P. Brouwer, L. Carretero-Paulet, S. Cheng, J. de Vries, P.-M. Delaux, A. Eily, N. Koppers, L.-Y. Kuo, Z. Li, M. Simenc, I. Small, E. Wafula, S. Angarita, M. S. Barker, A. Bräutigam, C. dePamphilis, S. Gould, P. S. Hosmani, Y.-M. Huang, B. Huettel, Y. Kato, X. Liu, S. Maere, R. McDowell, L. A. Mueller, K. G. J. Nierop, S. A. Rensing, T. Robison, C. J. Rothfels, E. M. Sigel, Y. Song, P. R. Timilsena, Y. Van de Peer, H. Wang, P. K. I. Wilhelmsson, P. G. Wolf, X. Xu, J. P. Der, H. Schluepmann, G. K.-S. Wong, K. M. Pryer, Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants*. **4**, 460–472 (2018).

99. S. Alaba, P. Piszczalka, H. Pietrykowska, A. M. Pacak, I. Sierocka, P. W. Nuc, K. Singh, P. Plewka, A. Sulkowska, A. Jarmolowski, W. M. Karlowski, Z. Szweykowska-Kulinska, The liverwort Pellia endiviifolia shares microtranscriptomic traits that are common to green algae and land plants. *New Phytol.* **206**, 352–367 (2015).

100. S. Gao, H.-N. Yu, Y.-F. Wu, X.-Y. Liu, A.-X. Cheng, H.-X. Lou, Cloning and functional characterization of a phenolic acid decarboxylase from the liverwort Conocephalum japonicum. *Biochem. Biophys. Res. Commun.* **481**, 239–244 (2016).

101. M. G. Johnson, C. Malley, B. Goffinet, A. J. Shaw, N. J. Wickett, A phylotranscriptomic analysis of gene family expansion and evolution in the largest order of pleurocarpous mosses (Hypnales, Bryophyta). *Mol. Phylogenet. Evol.* **98**, 29–40 (2016).

102. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

103. B. Haas, A. Papanicolaou, Others, TransDecoder (find coding regions within transcripts). *Github, nd https://github. com/TransDecoder/TransDecoder (accessed May 17, 2018)* (2015).

104. A. Bateman, E. Birney, R. Durbin, S. R. Eddy, R. D. Finn, E. L. Sonnhammer, Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260–262 (1999).

105. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. **22**, 1658–1659 (2006).

106. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. **28**, 3150–3152 (2012).

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

108. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

109. M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–12 (2006).

110. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. **25**, 1972–1973 (2009).

111. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).

112. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T. Lam, ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

113. G. Yu, T. T.-Y. Lam, H. Zhu, Y. Guan, Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).

114. T. Junier, E. M. Zdobnov, The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*. **26**, 1669–1670 (2010).

115. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).

116. Z. Zhang, J. Li, X.-Q. Zhao, J. Wang, G. K.-S. Wong, J. Yu, KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*. **4**, 259–263 (2006).

Acknowledgements:

The authors thank Ralph Quatrano, David Cove, and the Quatrano laboratory at Washington University in St. Louis for incubating the *C. purpureus* genome project; Bernie Hauser, Thomas Colquhoun, and Drake Garner for assisting with the hormone and light perturbations; Bernard Goffinet, Michelle Mack, Sharon Robinson, Todd Rosenstiel, Blanka Shaw, and the late Norton Miller for providing field collections. Susanne Renner, Sally Otto, Mark Kirkpatrick, and Matthew Hahn provided valuable feedback on a draft of the manuscript. The One Thousand Plant Transcriptome Initiative generously provided early access to moss and liverwort data, and the University of

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Florida Interdisciplinary Center for Biotechnology Research and HiPerGator provided vital technical support throughout the project. **Funding:** This work was supported by NSF DEB-1541005 and start-up funds from UF to SFM; microMORPH Cross-Disciplinary Training Grant, Sigma-Xi Grant-In-Aid of Research, and Society for the Study of Evolution Rosemary Grant Award to SBC; NSF DEB-1239992 to NJW; Emil Aaltonen Foundation and the University of Turku to SO; NSF DEB-1541506 to JGB and SFM. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. **Author contributions:** Conceptualization: SFM; Data curation: KB; Formal analysis: SBC, JJ, SS, JTL, FM, AS, GPT, NFP, SAR; Funding acquisition: SBC, SH, NJW, JGB, SAR, SFM; Investigation: SBC, JJ, ACP, SS, JTL, FM, AS, GPT, NFP, CC, MW, AL, CD, CS, JCM, REC, LMK, SO, SH, JBL, JGB, SAR, SFM; Methodology: SBC, ACP, JJ, JS, SFM; Project administration: KB, JG, JS, SFM; Resources: NJW, MJG, JG, JS, SFM; Supervision: MGJ, SAR, JG, JS, SFM; Visualization: SBC, JJ, JTL, AS, GPT; Writing – original draft: SBC, SFM; Writing – review and editing: SBC, ACP, JJ, JTL, FM, AS, GPT, CAS, LMK, JGB NJW, MGJ, SAR, JG, JS, SFM. All authors approve of the final draft of this manuscript. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** *Ceratodon purpureus* data for this manuscript can be found under NCBI BioProjects listed in Table S9. Genome assemblies and annotations can be found on Phytozome (https://phytozome-next.jgi.doe.gov/). *Brachythecium rivulare* data can be found under BioProject PRJNA639184. The Lanisha transposable element can be found in NCBI GenBank under MT647524. Published sequence data used in this study can be found in

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S23. Supporting documents for the phylogenomic analysis of sex-linked genes can be found on Dryad under https://doi.org/10.5061/dryad.v41ns1rsm. Requests for *C. purpureus* lines presented here should be addressed to stuartmcdaniel@ufl.edu.

List of Supplementary Materials:

Materials and Methods

Table S1-30

Fig S1-14

References (41 - 116)

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

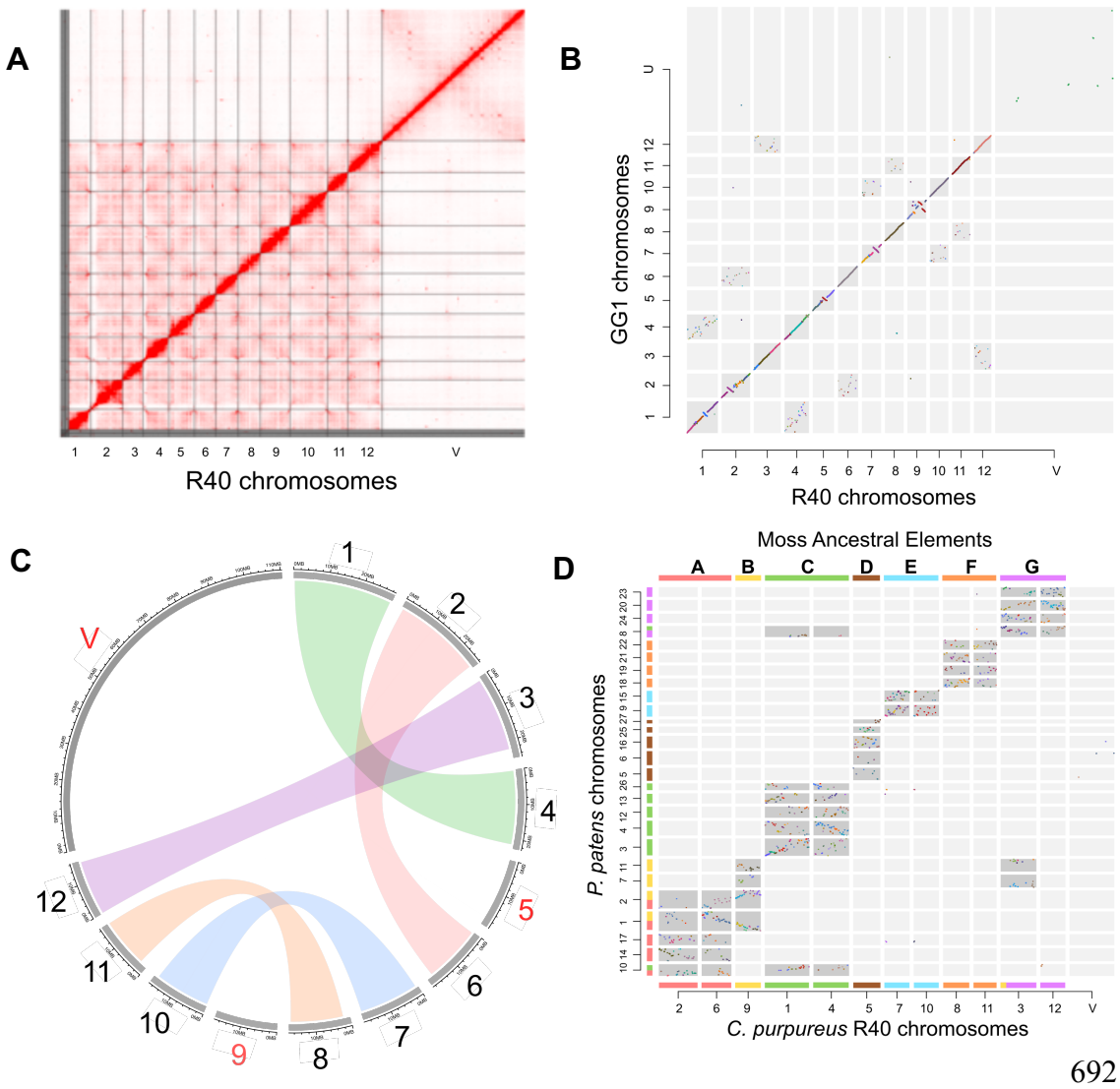

Fig. 1. Chromosome-scale genome of *C. purpureus* and synteny plots. A) Hi-C contact map of the *C. purpureus* R40 genome assembly. B) Dot plot between *C. purpureus* GG1 and R40 isolates. The U and V sex chromosomes lack a psuedoautosomal region. C) Self-synteny plot of *C. purpureus* R40 isolate; self-synteny of the GG1 isolate is similar (Fig. S3). D) Dot plot between *C. purpureus* R40 and *P. patens*, highlighting the moss ancestral elements (A-G).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

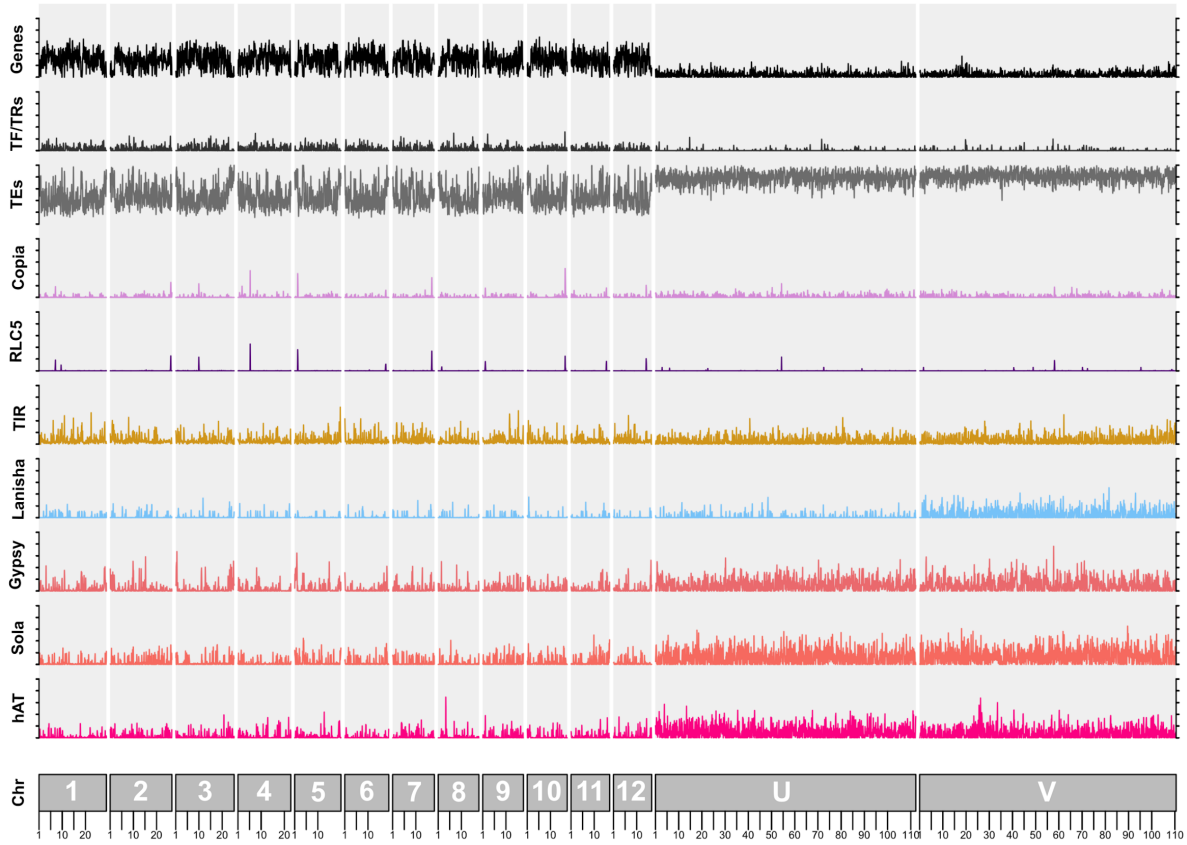

Fig. 2. Density plots across *C. purpureus* chromosomes (in Mb). Densities show the proportion (from 0 to 1) of a 100 Kilobase (Kb) window (90 Kb jump) containing each feature. The autosomes, U, and V sex chromosomes show different densities for certain features. However, local density peaks of RLC5 Copia elements on each chromosome represent candidate centromeric regions, similar to *P. patens* (*27*).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

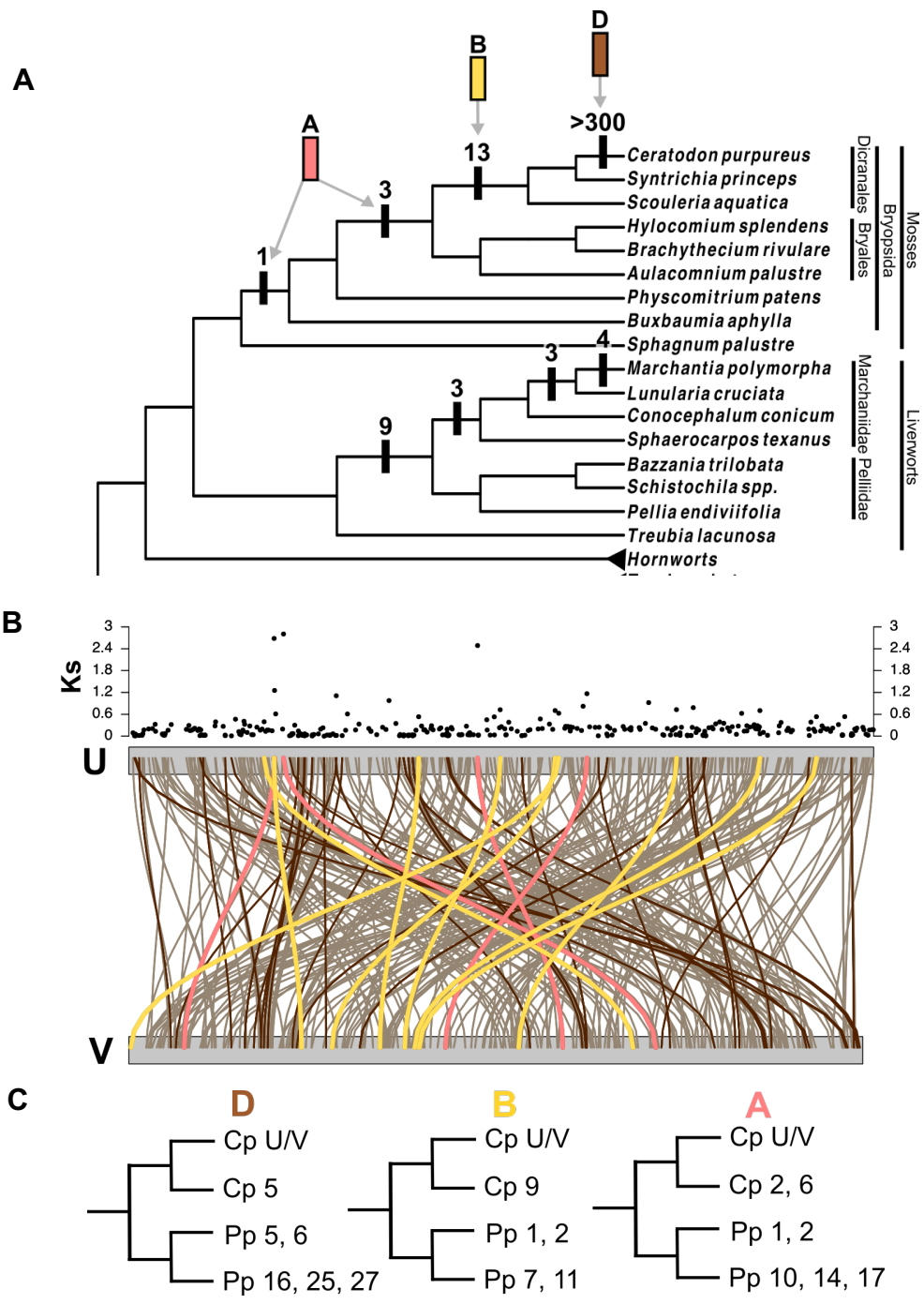

Fig. 3. Evolutionary history of moss and liverwort (Setaphyte) sex chromosomes. A) capture events across moss and liverwort sex chromosomes. Numbers indicate how many extant genes were captured based on the topology. The sex chromosome capture

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.03.163634; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

events can be traced back to three ancestral elements (A, B, and D). B) Ks of one-to-one U-V orthologs plotted on U and V sex chromosomes of *C. purpureus*. Lines connect one-to-one orthologs on the U and V, where colors correspond to capture events on the phylogeny: oldest genes, salmon; the third capture event, yellow; and the most recent, brown. The darker brown lines indicate genes with Ks ≤ 0.02, presumably the most recently captured genes. C) simplified gene trees showing the homeologous *P. patens* and *C. purpureus* chromosomes containing homologous genes, corresponding to ancestral elements D, B, and A.