

1 **No evidence that sex and transposable elements drive genome size**
2 **variation in evening primroses**

3 J Arvid Ågren^{1*}, Stephan Greiner^{2†}, Marc TJ Johnson^{3§}, and Stephen I Wright^{1‡}

4

5 ¹Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON,
6 M5S 3B2, Canada

7 ²Max Planck Institute of Molecular Plant Physiology, Wissenschaftspark Golm, Am
8 Mühlenberg 1, 14476 Potsdam–Golm, Germany

9 ³Department of Biology, University of Toronto–Mississauga, Mississauga, ON, L5L 1C6,
10 Canada

11

12 * arvid.agren@utoronto.ca, +1 416-978-8508 (corresponding author)

13 † greiner@mpimp-golm.mpg.de, +49 331-567-843; § marc.johnson@utoronto.ca, +1 905-
14 569-4484; ‡ stephen.wright@utoronto.ca, +1 416-946-8508

15

16 Running title: Sex and genome size in evening primroses

17 Key words: Mating system; functional asexuality; PTH; repetitive DNA; *Oenothera*

18 Word count: 3566; Abstract word count: 197

19 Number of figures: 2; Number of tables: 1; Number of supplementary tables: 2

20 Data archival location: C-value estimates will be submitted to Kew Garden's Plant DNA

21 C-Value Database (<http://data.kew.org/cvalues/>) and the paired-end sequences to the

22 Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>).

23

1 **Abstract**

2 Genome size varies dramatically across species, but despite an abundance of attention
3 there is little agreement on the relative contributions of selective and neutral processes in
4 governing this variation. The rate of sexual reproduction can potentially play an
5 important role in genome size evolution because of its effect on the efficacy of selection
6 and transmission of transposable elements. Here, we used a phylogenetic comparative
7 approach and whole genome sequencing to investigate the contribution of sex and
8 transposable element content to genome size variation in the evening primrose
9 (*Oenothera*) genus. We determined genome size using flow cytometry from 30
10 *Oenothera* species of varying reproductive system and find that variation in
11 sexual/asexual reproduction cannot explain the almost two-fold variation in genome size.
12 Moreover, using whole genome sequences of three species of varying genome sizes and
13 reproductive system, we found that genome size was not associated with transposable
14 element abundance; instead the larger genomes had a higher abundance of simple
15 sequence repeats. Although it has long been clear that sexual reproduction may affect
16 various aspects of genome evolution in general and transposable element evolution in
17 particular, it does not appear to have played a major role in the evening primroses.

18

19

20

21

22

23

1 **Introduction**

2 Variation in genome size is one of the most striking examples of biodiversity (Bennett
3 and Leitch 2012; Gregory 2013). Genomes may be as small as 160 kb, as in the obligate
4 endosymbiotic proteobacterium *Carsonella ruddii* (Nakabachi et al. 2006), and as large
5 as 150 GB in the polyploid plant *Paris japonica* (Pellicer et al. 2010). Variation is not
6 restricted to differences between species, as extensive genome size variation also exists
7 within species (Biemont 2008; Diez et al. 2013; Long et al. 2013). Understanding the
8 evolutionary processes underlying this variation has received much attention (reviewed in
9 for example Petrov 2001; Gregory 2005; Lynch 2007; Gaut and Ross-Ibarra 2008; Ågren
10 and Wright 2011), yet little consensus exists on the relative contributions of these
11 processes.

12 Variation in genome size may be influenced by both neutral and selective
13 evolutionary processes. Several studies have shown that genome size may evolve
14 neutrally, with increases and decreases mainly due to biases in insertion and deletion
15 rates and/or recombination rates (Petrov 2001; Oliver et al. 2007; Nam and Ellegren,
16 2012). It has also long been recognized that genome size correlates with various
17 ecologically relevant traits; with examples ranging from flowering time in plants
18 (Meagher and Vassiliadis, 2005) to powered flight in birds (Wright et al. 2014) and
19 selection may thus act adaptively on genome size. Much recent debate has focused on the
20 hypothesis that variation in the efficacy of selection, usually due to differences in
21 effective population size (N_e), governs most variation in genome size across distantly
22 related taxa (see for example exchanges by Lynch and Conery 2003; Charlesworth and
23 Barton 2004; Daubin and Moran 2004; Lynch and Conery 2004; Whitney et al. 2010,

1 Whitney and Garland 2010; Lynch 2011; Whitney et al. 2011). Finally, several studies
2 show that differential accumulation of transposable elements (TEs) can explain
3 differences in genome size (reviewed in Ågren and Wright 2011). Indeed, there appears
4 to be a correlation between relative TE abundance and genome size across angiosperms
5 (Tenailon et al. 2010). Furthermore, variation between closely related species, including
6 species of rice (Piegu et al. 2006), cotton (Hawkins et al. 2006), and *Arabidopsis* (Hu et
7 al. 2011), can be attributed to differences in TE abundance. Understanding what factors
8 allow TEs to accumulate in some species, but not in others, may therefore be central to
9 understanding variation in genome size in general, and among plants in particular.

10 A key factor affecting the efficacy of selection and TE transmission is the rate of
11 sexual reproduction (Hickey 1982; Wright and Schoen 1999; Morgan 2001; Docking et al.
12 2006; Dolgin and Charlesworth 2006; Glemin and Galtier 2012; Ågren 2014). Sexual
13 reproduction provides a way for TEs to spread to new lineages in a population, and
14 despite the deleterious effects on host fitness, theory predicts that TEs in sexual
15 populations should evolve maximum transposition rates (Charlesworth and Langley
16 1986). By contrast, within-lineage transmission of genes within asexual populations is
17 expected to limit TE spread and in the absence of horizontal gene transfer reduction in
18 transposition rate due to self-regulation is more likely to evolve in highly selfing and
19 asexual species (Charlesworth and Langley 1986). Taken together, this leads to the
20 prediction of higher TE abundances and larger genome sizes in sexually reproducing
21 species compared to asexually reproducing relatives. On the other hand, the reduced
22 effective population size in asexuals means that any actively transposing TEs they inherit
23 from sexual progenitors may initially accumulate to high numbers by a Muller's ratchet-

1 like process (Muller 1964; Gabriel et al. 1993; Charlesworth 2012). Over time, asexual
2 lineages with high TE abundance are expected to go extinct, whereas asexual lineages
3 with low TE abundance are more likely to persist (Arkhipova and Meselson 2005; Dolgin
4 and Charlesworth 2006). Whether sex will cause an increase or decrease in genome size
5 remains unclear and represents an important problem for understanding the forces that
6 affect genome evolution.

7 To date, most studies on the role of sexual reproduction on TE accumulation and
8 genome expansion have focused on animal systems (e.g Zeyl et al. 1996; Arkhipova and
9 Meselson 2000; Schaack et al. 2010) whereas work on plants has been lacking (but see
10 Docking et al. 2006). Studies on plants have typically compared self-fertilizing species to
11 outcrossing relatives (e.g Lockton and Gaut 2010; de la Chaux et al. 2012; Ågren et al.
12 2014). Several studies have shown an association between outcrossing rate and genome
13 expansion (Albach and Greilhuber 2004; Trivers et al. 2004; Wright et al. 2008; Hu et al.
14 2011), but it is difficult to distinguish whether these differences are due to TE
15 accumulation in the outcrossers or loss of TEs in the selfers (Wright and Ågren 2011).
16 Furthermore, when correcting for phylogenetic non-independence, Whitney et al. (2010)
17 failed to detect a significant effect of outcrossing rate on genome size in a broad scale
18 comparison of 205 species. Similar large-scale studies of the effect of sexual reproduction
19 on plant genome size evolution have so far been lacking.

20 The evening primrose plant family (Onagraceae) provides an ideal system to test
21 the effect of sexual reproduction on genome size evolution. Functional asexuality (i.e.
22 absence of recombination and segregation) in the Onagraceae has evolved more than 20
23 times independently due to a genetic system called Permanent Translocation

1 Heterozygosity (PTH), which is characterized by suppression of meiotic recombination
2 and segregation (Cleland 1972; Harte 1994; Rauwolf et al. 2008; Johnson et al. 2009).
3 Functional asexuality due to PTH has been described in eight plant families (Cleland
4 1972; Holsinger and Ellstrand, 1984; Harte 1994), and differs from apomixis in that
5 individuals go through all stages of meiosis, and successful zygote formation still requires
6 fertilization (see Whitton et al. 2008 for a review of other forms of plant asexuality).
7 Another useful contrast between PTH and many other asexual species is that PTH species
8 tend to share the same ploidy level with their sexual relatives, allowing the effect of
9 reproductive system to be decoupled from the effect of ploidy.

10 In this study we take a phylogenetic comparative approach to examine whether
11 the repeated transitions from sex to PTH in *Oenothera* has been associated with a shift in
12 genome size. We also use whole genome sequencing of three species of varying genome
13 size and reproductive system to assess the contribution of transposable elements to
14 genome size variation in the genus.

15

16 **Material and Methods**

17 *Study system*

18 In this study we focus on the *Oenothera* genus of the evening primrose family
19 Onagraceae. The genus is monophyletic (Levin et al. 2004; Wagner et al. 2007) and
20 includes the largest number of PTH species and their sexual relatives (Raven 1979;
21 Holsinger and Ellstrand 1984; Johnson et al. 2011). PTH in *Oenothera* results from three
22 mechanisms (Cleland 1972; Harte, 1994; Rauwolf et al. 2008). First, during metaphase I
23 of meiosis, all 14 chromosomes form a ring ($x = n = 7$) instead of bivalents, which

1 restricts synapsis and recombination to the highly homozygous telomeres, such that
2 genetically detectable recombination is effectively 0. Second, in anaphase I, alternate
3 disjunction results in one haploid set of chromosomes segregating as a unit, and the other
4 haploid set as another unit. Unless one of each unit is present in the zygote, the zygote
5 will not survive. This balanced mortality of gametes prevents segregation, which leads to
6 permanent heterozygosity. Finally, > 99.5% of seeds are self-fertilized (R Godfrey and
7 MTJ Johnson, unpublished results) because receptive stigmas accept pollen before
8 flowers open. In short, the genetics of PTH reproduction in *Oenothera* can be likened
9 with splitting the genome in half; only to later fuse the two halves back together, without
10 recombination or segregation (Cleland 1972; Harte, 1994; Rauwolf et al. 2008). Recent
11 evidence shows that recombination is also suppressed in sexual bivalent forming species,
12 suggesting that recombination rates may not be dramatically different between sexual and
13 PTH species. However, sexual species still undergo free segregation of homologous
14 chromosomes, which should effectively eliminate genetic linkage disequilibria between
15 chromosomes and allow the formation of homozygous loci from heterozygous parents as
16 per Hardy-Weinberg expectations (Rauwolf et al. 2011; Golczyk et al. 2014). By contrast,
17 these processes are completely lost in PTH species, leading to the perpetual propagation
18 of single genotypes (Stebbins 1950; Cleland 1972).

19

20 *Genome size estimates*

21 In this study we examined genome size in 30 *Oenothera* species, including 17 PTH and
22 13 sexual species (Figure 1). Sterilized seeds were stratified by sowing them on agar and
23 kept in the dark at 4°C for three weeks. Seedlings were transferred to pots and grown in

1 controlled glasshouse conditions between October 2013 and January 2014 at the
2 University of Toronto. Leaf tissue, sampled at the same level of maturity (~ 5 weeks),
3 from each species was sent to Plant Cytometry Services (PO Box 299, 5480 AG
4 Schijndel, The Netherlands), who determined DNA content in picograms (pg) using flow
5 cytometry with Propidium Iodide fluorescent dye and *Pachysandra terminalis* as a
6 standard (1C = 1.73; Zonneveld et al. 2005). This method has previously been shown to
7 successfully detect small differences in genome size, including within species variation
8 (Diez et al. 2013). Three replicates were performed per species.

9

10 *Phylogenetic analysis*

11 To control for statistical non-independence due to shared evolutionary history
12 (Felsenstein 1985), we accounted for phylogeny in our statistical analysis. We inferred
13 the phylogeny of the 30 species using the previously generated phylogeny of *Oenothera*
14 in Johnson et al. (2009). Briefly, we sequenced two plastid (*trnL-trnF* and *rps16*) and
15 three nuclear gene regions (*PgiC*, *ITS*, and *ETS*) from 121 species and created a
16 maximum likelihood phylogeny using RAxML 7.0.4 (Stamatakis 2006). The tree was
17 made ultrametric using non-parametric rate smoothing in TreeEdit
18 (<http://tree.bio.ed.ac.uk/software/treededit>) and then pruned to include only the 30 species
19 studied here. For further details of the phylogeny see Johnson et al. (2009).

20 To start, we tested whether the data exhibited significant phylogenetic signal
21 using Pagel's λ (Pagel 1999) implemented in the *phylosig* function in the *phytools*
22 package (Revell 2012) of R version 3.0.2 (R Development Core Team 2013). This
23 package assesses the significance of phylogenetic signal by performing a likelihood ratio

1 test against the null hypothesis that $\lambda = 0$. Next, we performed phylogenetic generalized
2 least squares (PGLS; Butler and King 2004) regression between reproductive system (sex
3 was coded as 0 and PTH as 1) and genome size using the *ape* (Paradis et al. 2004) and
4 *geiger* (Harmon et al. 2008) packages in R. We performed the PGLS tests under both
5 neutral (Brownian motion) and stabilizing selection (Ornstein-Uhlenbeck) models of trait
6 evolution. Akaike's Information Criterion (AIC) was used to determine which model best
7 described the data.

8

9 *DNA isolation and Illumina sequencing*

10 Seeds of *O. elata* (sex) and *O. biennis* (PTH), two species with relatively large genome
11 sizes, and of *O. villaricae* (PTH), one of relatively small genome size (Supplementary
12 Table 1), were imbibed at 4°C overnight in the dark with 3% Plant Preservative Mixture
13 (Plant Cell Technology Store, Washington, DC, USA). Seeds were then sown on the
14 substrate surface and grown until the early rosette stage (for details see Greiner and Köhl,
15 2014). For DNA isolation, 50 mg of leaf material was frozen in liquid N₂ and ground
16 using a mixer mill. After adding 775 µl of IGEPAL-buffer [1.9% IGEPAL, 1.9% CTAB,
17 130 mM Tris/HCl (pH 8.0), 130 mM EDTA (pH 8.0), 1.3% PVP-40, 1.8 M NaCl, 130
18 mM B(OH)₃], 125 µl 1-thioglycerol, and 100 µl RNase A (50 mg/ml), samples were
19 incubated at 60°C for 30 minutes under medium shaking. Cell debris was removed by
20 centrifugation. The supernatant was treated twice with chloroform/isoamylalcohol (24:1)
21 and subsequently with phenol/chloroform/isoamylalcohol (25:24:1, pH 7.5). The
22 phenol/chloroform/isoamylalcohol treatment was repeated until the aqueous phase was no
23 longer “reddish”. DNA was precipitated with 1/10 volume of 5 M NH₄-acetate and 1

1 volume of isopropanol (-20°C), incubated at -20°C for 15 min, collected by
2 centrifugation, and washed with 70% and/100% EtOH (-20° C). Resolved DNA was
3 further purified using the Genomic DNA Clean & Concentrator™ kit (Zymo Research
4 Cooperation, Irvine, CA, USA). DNA was sequenced at the Max Planck-Genome-Centre
5 Cologne (Germany) on an Illumina HiSeq2500 platform, 100 bp paired-end, utilizing a
6 TruSeq DNA library (375 bp insert size). All sequences will be deposited to the Sequence
7 Read Archive (<http://www.ncbi.nlm.nih.gov/sra>).

8

9 *Repetitive content analysis*

10 To assess whether variation in genome size could be attributed to differential
11 accumulation of repetitive elements such as TEs, we determined the repetitive content in
12 three species of varying genome size and reproductive system. To characterize repeats we
13 ran RepeatExplorer (Novak et al. 2013), a *de novo* graph-clustering pipeline for repeat
14 characterization (Novak et al. 2010), which is implemented in the Galaxy platform
15 (<http://galaxyproject.org>). We filtered reads for quality, keeping only reads with a Phred
16 quality score of at least 20 over 90% of their length. RepeatExplorer joins reads together
17 in clusters based on sequence similarity and then matches these clusters against RepBase
18 (Jurka et al. 2005) to identify repeats. We ran the pipeline on one sample per species
19 under default settings (Table 1).

20

21 **Results**

22 *Variation in genome size*

23 We detected almost two-fold variation in genome size among the diploid species

1 surveyed (Figure 1; Supplementary Table 2). Estimated genome size ranged from 0.64 pg
2 in *O. mendocinensis* (sex) and *O. sandiana* (PTH), to 1.16 pg in *O. nutans* (PTH). The
3 South American clade (Subsection *Munzia*: *O. mendocinensis*, *O. nana*, etc.) consistently
4 had the smallest genomes, whereas the recently radiated North American *O. biennis* clade
5 (Subsection *Oenothera*) consistently had the largest genome sizes. The so-called “B clade”
6 (sensu Wagner et al. 2007; e.g. *O. flava*, *O. tetraptera*, *O. perennis*, *O. fruticosa*, *O. rosea*,
7 *O. kunthiana*) varied the most in genome size, reflecting polyploidy (exact ploidy level
8 unknown) in one species (*O. fruticosa*) and deep divergence among multiple lineages.
9 Excluding the polyploid *O. fruticosa* (sex), the mean genome size was 0.85 +/- 0.036 s.e.
10 pg. C-value estimates will be submitted to Kew Garden’s Plant DNA C-Value Database
11 (<http://data.kew.org/cvalues/>)

12

13 *Phylogenetic signal*

14 We detected significant phylogenetic signal in genome size across the species examined.
15 Pagel’s λ was 0.74 ($P = 0.0141$ that $\lambda > 0$) suggesting that phylogenetic non-
16 independence should be taken into account in statistical analyses.

17

18 *Phylogenetic generalized least squares analysis*

19 We found no significant relationship between sexual reproduction and genome size. The
20 lack of an effect of sex on genome size holds regardless of whether we assume that
21 genome size evolves under a neutral Brownian motion model (df = 29, $P = 0.828$) or
22 moving towards a selective optimum in an Ornstein-Uhlenbeck’s model of stabilizing
23 selection (df = 29, $P = 0.8162$). Comparing the AIC scores suggests that the Brownian

1 motion model (AIC = -103.4203) better describes the data than the Ornstein-Uhlenbeck
2 model (AIC = -42.8108).

3

4 *Repetitive content*

5 Repetitive elements were abundant in all three species examined. After filtering for
6 quality, about 80% of all reads formed clusters (Table 1). In all three species TEs made
7 up most of the repetitive content, with the dominant TEs being long terminal repeat *gypsy*
8 and *copia* elements (Figure 2). In all three species, we estimated that TEs make up ~ 35-
9 40% of the genome.

10 To investigate whether the genomes differed in other kinds of repeats, we
11 combined the sequences annotated as “simple repeat”, “satellite”, and “low complexity”
12 under the label “short simple repeats”. These repeats made up a larger proportion in the
13 relatively large genomes *O. elata* and *O. biennis* (~ 35%), than in the smaller *O.*
14 *villaricae* (~ 20%; Figure 2) and this difference was statistically significant (Pearson’s
15 chi-square test of independence, $\chi^2 = 59360.79$, $df = 2$, $P = 2.2 \times 10^{-16}$).

16

17 **Discussion**

18 Studies of genome size variation have a long history (Mirsky and Ris 1951). References
19 to this variation have featured heavily in the arguments about the role of non-selective
20 processes in the evolution of genome complexity (Lynch and Conery 2003; Lynch 2007;
21 Whitney and Garland 2010; Whitney et al. 2010; Lynch 2011; Whitney et al. 2011) and
22 more recently in the debate associated with the ENCODE Project Consortium’s claim
23 that 80% of the human genome can be assigned a biochemical function (ENCODE

1 Project Consortium 2012; Eddy 2012; Graur et al. 2013; Doolittle 2013; Kellis et al.
2 2014; Palazzo and Gregory 2014). With estimates from some 7,000 species and a 2,400-
3 fold variation in genome size (Leitch and Leitch 2013), studies of plant genomes have
4 much to contribute to these and other debates.

5 Here, we presented genome size estimates in thirty species in the evening
6 primrose genus *Oenothera* and found no evidence that sex explains the almost two-fold
7 variation in genome size. Instead, evolution of genome size was fairly conserved within
8 *Oenothera* and best explained by neutral genetic drift, as opposed to a model of
9 stabilizing selection towards an optimum, or a model that ignores evolutionary history.
10 Moreover, contrary to the reasoning outlined in the Introduction, we found no evidence
11 that genome size variation in *Oenothera* can be attributed to transposable element
12 abundance. Instead the observable difference in genome size appeared to be due to
13 accumulation of short simple repeats.

14 Central to the hypothesis that sexual species should have larger genomes than
15 asexual relatives is that sexual reproduction should facilitate the spread of transposable
16 elements (Hickey 1982; Charlesworth and Langley 1986; Morgan 2001; Dolgin and
17 Charlesworth 2006). Our results are in line with two recent studies highlighting that
18 differences in TE abundance are not the only reason why lineages may differ in genome
19 size. For example, in *Eucalyptus*, a member of the same order as *Oenothera* (Myrtales), a
20 difference in abundance of tandem repeats rather than TEs is responsible for the 110 Mb
21 difference in genome size between the closely related *E. grandis* and *E. globulus*
22 (Myburg et al. 2014). In *Arabidopsis*, the highly selfing *A. thaliana* has fewer TEs and
23 smaller genome than its outcrossing relative *A. lyrata* (Hu et al. 2011), which is likely

1 due to accumulation in the outcrosser rather than a loss of TEs in the selfer (Slotte et al.
2 2013), consistent with TE transmission advantage in an outcrossing lineage. However,
3 genome size variation within *A. thaliana* does not seem to be due to TEs. In particular, Long
4 et al. (2013) detected a 10% variation in genome size among 180 Swedish *Arabidopsis*
5 *thaliana* lines, which was due to differential accumulation of 45S rDNA rather than TEs.
6 What determines whether genome size difference will be due to TEs or simple repeats
7 remains unclear.

8 Although there is evidence from multiple systems that sex may promote the
9 spread of TEs (Zeyl et al. 1996; Arkhipova and Meselson 2000; Schaack et al. 2010),
10 there is also abundant evidence that asexuality is associated with a reduction in the
11 efficacy of selection (reviewed in for example Glemin and Galtier 2012). Consistent with
12 this, a recent study of 13 sexual and 16 PTH *Oenothera* species found evidence of a
13 reduction in the efficacy of selection in PTH species (Hollister et al. submitted). If the
14 magnitude of this reduction in the efficacy of selection is large enough, it may dilute the
15 transmission advantage associated with sex. This kind of dilution effect could be
16 magnified if the sexual species vary in selfing rate. However, variation in selection
17 efficacy due to variation in selfing rate is unlikely to explain our results because all
18 sexual species examined, except *O. versicolor* and *O. mendocinensis*, are either self-
19 incompatible or highly herkogamous, i.e. having spatially separated anthers and stigma.

20 The results from our within-genus comparison corroborate those of the multi-
21 family study of the role of outcrossing rate in genome size evolution by Whitney et al.
22 (2010). In their paper, they find no effect of outcrossing rate on genome size and suggest
23 that an effect of mating system in their analysis could have been obscured by rapid

1 mating system shifts. In our analysis, the short time scale due the recent divergence
2 between species may mean that any effect of reproductive system have yet to materialize.
3 Using computer simulations to examine changes in TE copy number following a shift to
4 asexual reproduction, Docking et al. (2006) found that it took around 50,000 generations
5 to reach a new equilibrium. During the first 30,000 generations following the shift to
6 asexuality, copy numbers often increased dramatically, but even with low levels of
7 excisions all elements were lost. At any given point in time, however, variation among
8 lineages was very high. Thus although quantitative predictions from computer
9 simulations will be sensitive to specific parameter values assumed, they do highlight that
10 sampling a limited number of lineages could fail to detect a correlation between sex and
11 TE copy number, especially if divergence times are small such is the case in *Oenothera*.
12 If the rapid shifts of mating and reproductive systems mean that the effect on TE and
13 genome size evolution on both short (within genus) or long (multi-family) time scales is
14 difficult to capture, a possible middle road could be to examine within family variation.
15 The plant family Brassicaceae, home of *Arabidopsis*, is a good candidate for such a
16 family. Within the Brassicaceae, genome size varies 16-fold across the almost two
17 hundred species (Lysak et al. 2009; Johnston et al. 2005), species vary extensively in
18 mating system (Vekemans et al. 2014), and the large number of sequenced reference
19 genomes (Haudry et al. 2013) make it suitable for comparative genomic analysis.

20 Here, we have performed the first comprehensive study of the role of sex in
21 genome size evolution in plants. Whereas it has long been clear that sexual reproduction
22 affects various aspects of genome evolution, including in the evening primroses

1 (Ellstrand and Levin 1980; Hersch-Green et al. 2012; Hollister et al. submitted), it
2 appears to have played no role in genome size evolution in *Oenothera*.

3

4 **Acknowledgements**

5 We thank Niroshini Epitawalage and Victor Mollov for assistance in the glasshouse,
6 Brian Husband and Paul Kron for discussion and advice on flow cytometry, and Elena
7 Ulbricht-Jones and Liliya Yaneva-Roder for providing the DNA samples for Illumina
8 sequencing. Liliya Yaneva-Roder is further acknowledged for excellent technical
9 assistance during the establishment of the phenol-based DNA isolation protocol for
10 *Oenothera*. Seeds were generously donated by USDA GRIN, Algonquin Park Visitor
11 Centre staff, S.C.H. Barrett, Ornamental Plant Germplasm Centre (OPGC), J. Meurer, W.
12 Wagner, and R. Raguso. J.A.Å is supported by a Junior Fellowship from Massey College,
13 S.G. by the Max Planck Society and the Deutsche Forschungsgemeinschaft (DFG),
14 M.T.J.J. and S.I.W. by the Natural Sciences and Engineering Research Council (NSERC).
15 We declare no conflict of interest.

16

17

18

19

20

21

1 FIGURE 1 Molecular phylogeny of *Oenothera* used in the PGLS analysis. Haploid
2 genome size estimates (pg; 1 pg = 978 Mb) are plotted on top. Note, *O. fruticosa* is a
3 polyploid and not included in the statistical analyses.

4

5 FIGURE 2 Repetitive content of three *Oenothera* species as estimated by RepeatExplorer
6 (Novak et al. 2013), including two species of relatively large genome size (*O. elata* and *O.*
7 *biennis*) and one of relatively small genome size (*O. villaricae*).

8

9 **Table 1 Clustering statistics for RepeatExplorer**

	<i>O. villaricae</i>	<i>O. elata</i>	<i>O. biennis</i>
Total number of reads used in analysis	1001571	1012981	729978
Number of reads in clusters	813712	857862	639924
Number of clusters	24496	19474	12197

10

11

12

13

14

15

16

17

18

19

20

21

1 **References**

2 Ågren, J. A., and S. I. Wright. 2011. Co-evolution between transposable elements and
3 their hosts: a major factor in genome size evolution? *Chromosome. Res.* 19:777–786.

4

5 Ågren, J. A. 2014. Evolutionary transitions in individuality: insights from transposable
6 elements. *Trends Ecol. Evol.* 29:90–96.

7

8 Ågren, J. A., W. Wang, D. Koenig, B. Neuffer, D. Weigel, and S. I. Wright. 2014.
9 Mating system shifts and transposable element evolution in the plant genus *Capsella*.

10 BMC Genomics In press.

11

12 Albach, D. C., and J. Greilhuber. 2004. Genome size variation and evolution in *Veronica*.
13 *Ann. Bot.* 94:897–911.

14

15 Arkhipova, I., and M. Meselson. 2000. Transposable elements in sexual and ancient
16 asexual taxa. *Proc. Natl. Acad. Sci. USA* 97:14473–7.

17

18 Arkhipova, I., and M. Meselson. 2005. Deleterious transposable elements and the
19 extinction of asexuals. *BioEssays* 27:76–85.

20

21 Bennett, M. D., and I. J. Leitch. 2012. Plant DNA C-values Database.

22 <http://data.kew.org/cvalues/>.

23

- 1 ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genom
2
- 3 Biémont, C. 2008. Within-species variation in genome size. *Heredity (Edinb)*. 101:297–
4 298.
5
- 6 Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling
7 approach for adaptive evolution. *Am. Nat.* 164:683–695.
8
- 9 Charlesworth, B., and C.H. Langley. 1986. The evolution of self-regulated transposition
10 of transposable elements. *Genetics* 112:359–383.
11
- 12 Charlesworth, B., and N. Barton. 2004. Genome size: does bigger mean worse? *Curr.*
13 *Biol.* 14:R233–235
14
- 15 Charlesworth, B. 2012. The effects of deleterious mutations on evolution at linked sites.
16 *Genetics* 190:5–22.
17
- 18 Cleland, R. E. 1972. *Oenothera: Cytogenetics and Evolution*. Academic Press, New
19 York, NY, USA.
20
- 21 Daubin, V., and N. A. Moran. 2004. Comment on "The Origins of Genome Complexity".
22 *Science* 306:978.
23

- 1 de la Chaux, N., T. Tsuchimatsu, K. S. Shimizu, and A. Wagner. 2012. The
2 predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in
3 transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*.
4 *Mobile DNA* 3:2
5
6 Diez, C. M., B. S. Gaut, E. Meca, E. Scheinvar, S. Montes-Hernandez, L. E. Eguiarte,
7 and M. I. Tenailon. 2013. Genome size variation in wild and cultivated maize along
8 altitudinal gradients. *New Phytol.* 199:264–276.
9
10 Docking, T. R., F. E. Saade, M. C. Elliott, and D. J. Schoen. 2006. Retrotransposon
11 sequence variation in four asexual plant species. *J. Mol. Evol.* 62:375–387.
12
13 Dolgin, E. S., and B. Charlesworth. 2006. The fate of transposable elements in asexual
14 populations. *Genetics* 174:817–827.
15
16 Doolittle, W. F. 2013. Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci.*
17 *USA* 110:5294–5300.
18
19 Eddy, S. R. 2012. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* 22: R898–
20 899.
21
22 Ellstrand, N. C., and D. A. Levin. 1980. Recombination system and population structure
23 in *Oenothera*. *Evolution* 34:923–933.

1

2 Felsenstein, J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.

3

4 Ford, V. S., and L. D. Gottlieb. 2007. Tribal relationships within Onagraceae inferred
5 from *PgiC* sequences. *Syst. Botany* 32:348–356.

6

7 Gabriel, W., M. Lynch, and R. Bürger. 1993. Muller's ratchet and mutational meltdowns.
8 *Evolution* 47:1744–1757.

9

10 Gaut, B. S., and J. Ross-Ibarra. 2008. Selection on major components of angiosperm
11 genomes. *Science* 320:484–486.

12

13 Golczyk, H., A. Massouh, and S. Greiner. 2014. Translocations of chromosome end-
14 segments and facultative heterochromatin promote meiotic ring formation in evening
15 primroses. *Plant Cell* 26:1280–1293.

16

17 Graur, D., Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall, and E. Elhaik. 2013. On the
18 immortality of television sets: "function" in the human genome according to the
19 evolution-free gospel of ENCODE. *Genome Biol. Evol.* 5:578–590.

20

21 Gregory, T. R. 2005. *The Evolution of the Genome*. Elsevier Academic, Burlington, MA,
22 USA.

23

- 1 Gregory, T. R. 2013. Animal Genome Size Database. <http://www.genomesize.com/>.
- 2
- 3 Greiner, S., and K. Köhl. 2014. Growing evening primroses (*Oenothera*). *Front. Plant*
- 4 *Sci.*5:38.
- 5
- 6 Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER:
- 7 investigating evolutionary radiations. *Bioinformatics* 24:129–131.
- 8
- 9 Harte, C. 1994. *Oenothera: Contributions of a Plant to Biology*. Springer-Verlag, New
- 10 York, NY, USA.
- 11
- 12 Haudry, A., A. E. Platts, E. Vello, D. R. Hoen, M. Leclercq, R. J. Williamson, E.
- 13 Forczek, Z. Joly-Lopez, J. G. Steffen, K. M. Hazzouri et al. 2013. An atlas of over 90,000
- 14 conserved noncoding sequences provides insight into crucifer regulatory regions. *Nat.*
- 15 *Genet.* 45:891–898.
- 16
- 17 Hawkins, J. S., H. Kim, J. D. Nason, R. A. Wing, and J. F. Wendel. 2006. Differential
- 18 lineage-specific amplification of transposable elements is responsible for genome size
- 19 variation in *Gossypium*. *Genome Res.* 16:1252–61.
- 20
- 21 Hersch-Green, E. I., H. Myburg, and M. T. J. Johnson. 2012. Adaptive molecular
- 22 evolution of a defence gene in sexual but not functionally asexual evening primroses. *J.*
- 23 *Evol. Biol.* 25:1576–1586.

1

2 Hickey, D. A. 1982. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*
3 101:519–531.

4

5 Holsinger, K. E., and N. C. Ellstrand. 1984. The evolution and ecology of permanent
6 translocation heterozygotes. *Am. Nat.* 124: 48–71.

7

8 Hu, T.T., P. Pattyn, E. G. Bakker, J. Cao, J. F. Cheng, R. M. Clark, N. Fahlgren, J. A.
9 Fawcett, J. Grimwood, H. Gundlach et al. 2011. The *Arabidopsis lyrata* genome
10 sequence and the basis of rapid genome size change. *Nat. Genet.* 43:476–481.

11

12 Johnson, M. T. J., S. D. Smith, and M. D. Rausher. 2009. Plant sex and the evolution of
13 plant defenses against herbivores. *Proc. Natl. Acad. Sci. USA* 106: 18079–18084.

14

15 Johnson, M.T.J, R. FitzJohn, S. D. Smith, M. D. Rausher, and S.P. Otto. 2011. Loss of
16 sexual recombination and segregation is associated with increased diversification in
17 evening primroses. *Evolution* 65:3230–3240.

18

19 Johnston, J. S., A. E. Pepper, A. E. Hall, Z. J. Chen, G. Hodnett, J. Drabek, R. Lopez, and
20 H. J. Price. 2005. Evolution of genome size in Brassicaceae. *Ann. Bot.* 95: 229–235.

21

22 Jurka, J, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz. 2005.
23 Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.*

1 110:462–467.
2
3 Kellis, M., B. Wold, M. P. Snyder, B. E. Bernstein, A. Kundaje, G. K. Marinov, L. D.
4 Ward, E. Birney, G. E. Crawford, J. Dekker, et al. 2014. Defining functional DNA
5 elements in the human genome. Proc. Natl. Acad. Sci. USA 111:6131–6138.
6
7 Leitch, I. J., and A. R. Leitch. 2013. Genome size diversity and evolution in land plants.
8 In: Plant Genome Diversity, Volume 2, (Leitch, I. J., Greilhuber, J., Dolezel, J. and
9 Wendel, J. F., eds.). Springer-Verlag, Vienna, Austria.
10
11 Levin, R. A., W. L. Wagner, P. C. Hoch, W. J. Hahn, A. Rodriguez, D. A. Baum, L.
12 Katinas, et al. 2004. Paraphyly in tribe Onagreae: insights into phylogenetic relationships
13 of Onagraceae based on nuclear and chloroplast sequence data. Syst. Botany 29:147–164.
14
15 Lockton, S, and B. S. Gaut. 2010. The evolution of transposable elements in natural
16 populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative
17 *Arabidopsis lyrata*. BMC Evol. Biol. 10: 10.
18
19 Long, Q., F. A. Rabanal, D. Meng, C. D. Huber, A. Farlow, A. Platzer, Q. Zhang, B. J
20 Vilhjálmsson, A. Korte, V. Nizhynska, et al. 2013. Massive genomic variation and
21 strong selection in *Arabidopsis thaliana* lines from Sweden. Nat. Genet. 45:884–890.
22
23 Lynch, M. and J. S. Conery. 2003. The origins of genome complexity. Science 302:1401–

1 1404.

2

3 Lynch, M., and J. S. Conery. 2004. Response to Comment on "The Origins of Genome
4 Complexity". *Science* 306:978

5

6 Lynch, M. 2007. *The Origins of Genome Architecture*. Sinauer Associates, Sunderland,
7 MA, USA.

8

9 Lynch, M. 2011. Statistical inference on the mechanisms of genome evolution. *PLoS*
10 *Genet.* 7.

11

12 Lysak, M. A., M. A. Koch, J. M. Beaulieu, A. Meister, and I. J. Leitch. 2009. The
13 dynamic ups and downs of genome size evolution in Brassicaceae. *Mol. Biol. Evol.*
14 26:85–98.

15

16 Meagher, T. R., and C. Vassiliadis. 2005. Phenotypic impacts of repetitive DNA in
17 flowering plants. *New Phytol.* 168:71–80.

18

19 Mirsky, A. E., and H. Ris. 1951. The desoxyribonucleic acid content of animal cells and
20 its evolutionary significance. *J. Gen. Physiol.* 34:451–462.

21

22 Morgan, M. T. 2001. Transposable element number in mixed mating populations. *Genet.*
23 *Res.* 77:261–275.

1

2 Muller, H.J. 1964. The relation of recombination to mutational advance. *Mutat. Res.* 1:2–

3 9.

4

5 Myburg, A.A., D. Grattapaglia, G. A. Tuskan, U. Hellsten, R. D. Hayes, J. Grimwood, J.

6 Jenkins, E. Lindquist, H. Tice, D. Bauer, et al. 2014. The genome of *Eucalyptus grandis*.

7 *Nature* 509:356-362

8

9 Nakabachi, A., A. Yamashita, H. Toh, H. Ishikawa, H. E. Dunbar, N. A. Moran, and M.

10 Hattori. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*.

11 *Science* 314:267.

12

13 Nam, K., and H. Ellegren. 2012. Recombination drives vertebrate genome contraction.

14 *PLoS Genet.* 8:e1002680.

15

16 Novak, P., P. Neumann, and J. Macas. 2010. Graph-based clustering and characterization

17 of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11:378.

18

19 Novak, P., P. Neumann, J. Pech, J. Steinhaisl, and J. Macas. 2013. RepeatExplorer: a

20 Galaxy-based web server for genome-wide characterization of eukaryotic repetitive

21 elements from next-generation sequence reads. *Bioinformatics* 29:792–793.

22

23 Oliver, M. J., D. Petrov, D. Ackerly, P. Falkowski, and O. M. Schofield. 2007. The mode

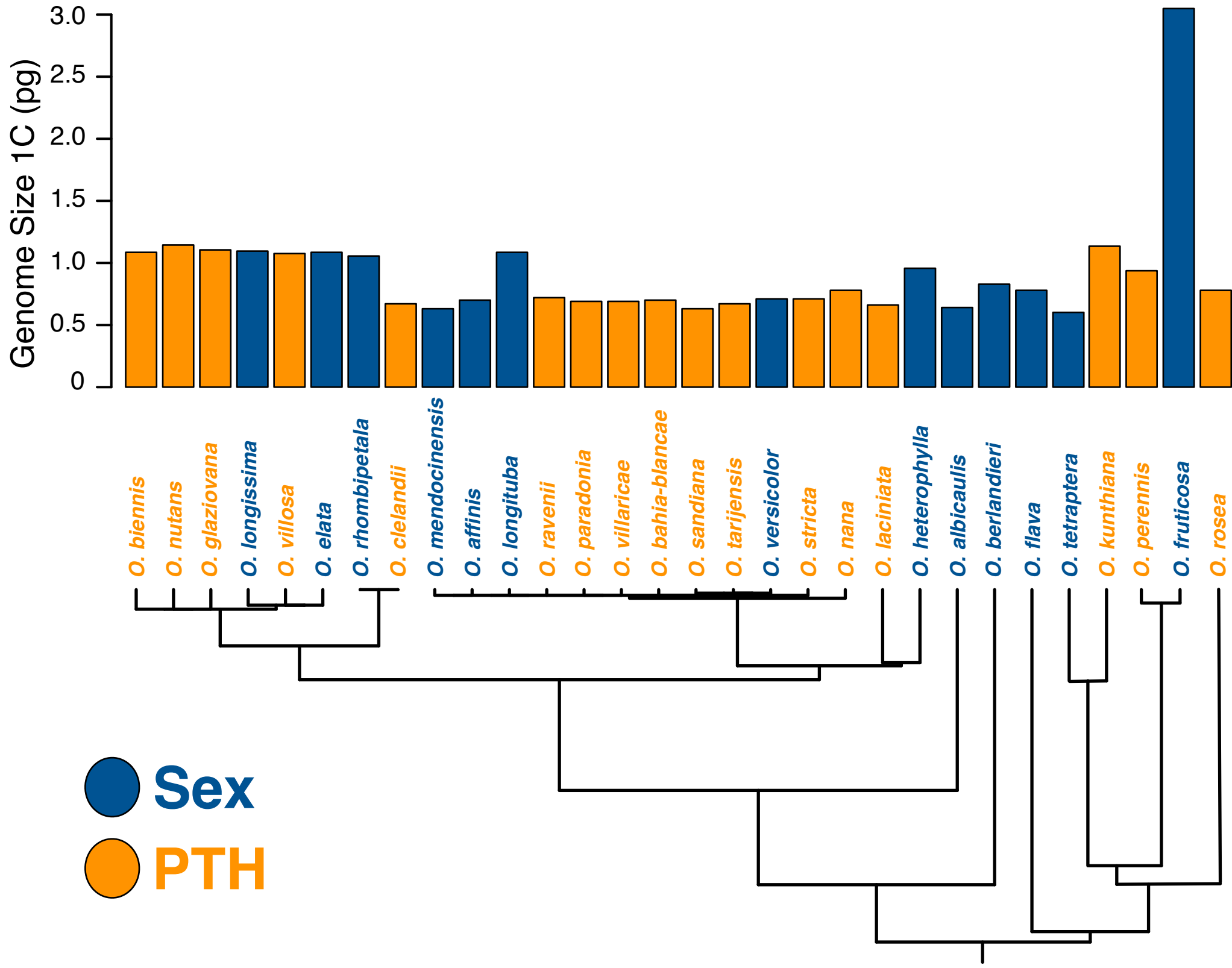
- 1 and tempo of genome size evolution in eukaryotes. *Genome Res.* 17:594-601.
- 2
- 3 Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–
- 4 884.
- 5
- 6 Palazzo, A. F., and T. R. Gregory. 2014. The case for junk DNA. *PLoS Genet*
- 7 10:e1004351.
- 8
- 9 Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and
- 10 evolution in R language. *Bioinformatics* 20:289–290.
- 11
- 12 Pellicer, J., M. F. Fay, and I. J. Leitch. 2010. The largest eukaryotic genome of them all?
- 13 *Bot. J. Linn. Soc.* 164:10–15.
- 14
- 15 Petrov, D. A. 2001. Evolution of genome size: new approaches to an old problem. *Trends*
- 16 *Genet.* 17:23–28.
- 17
- 18 Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Sanyal, H. Kim, K. Collura, D. S. Brar, S.
- 19 Jackson, R. A. Wing, and O. Panaud. 2006. Doubling genome size without
- 20 polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza*
- 21 *australiensis*, a wild relative of rice. *Genome Res* 16:1262–1269.
- 22
- 23 R Development Core Team. 2013. R: A language and environment for statistical

- 1 computing. R Foundation for Statistical Computing, Vienna, Austria.
- 2
- 3 Rauwolf, U., H. Golczyk, J. Meurer, R. G. Herrmann, and S. Greiner. 2008. Molecular
4 marker systems for *Oenothera* genetics. *Genetics* 180:1289–1306.
- 5
- 6 Rauwolf, U., S. Greiner, J. Mráček, M. Rauwolf, H. Golczyk, V. Mohler, R. G.
7 Herrmann, and J. Meurer. 2011. Uncoupling of sexual reproduction from homologous
8 recombination in homozygous *Oenothera* species. *Heredity (Edinb)*. 107:87–94.
- 9
- 10 Raven, P. H. 1979. A survey of reproductive-biology in Onagraceae. *New Zeal. J. Bot.*
11 17:575–593.
- 12
- 13 Revell, L. J., L. J. Harmon, and D. C. Collar. 2008. Phylogenetic signal, evolutionary
14 process, and rate. *Syst. Biol.* 57:591–601.
- 15
- 16 Schaack, S., E. J. Pritham, A. Wolf, and M. Lynch. 2010. DNA transposon dynamics in
17 populations of *Daphnia pulex* with and without sex. *Proc. R. Soc. B* 277:2381–2387.
- 18
- 19 Slotte, T., K. M. Hazzouri, J. A. Ågren, D. Koenig, F. Maumus, Y. L. Guo, K. Steige, A.
20 E. Platts, J. S. Escobar, L. K. Newman, et al. 2013. The *Capsella rubella* genome and the
21 genomic consequences of rapid mating system evolution. *Nat. Genet.* 45:831–5.
- 22
- 23 Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic

- 1 analyses with thousands of taxa and mixed models. *Bioinformatics*. 22:2688–2690.
- 2
- 3 Stebbins, G. L. 1950. *Variation and Evolution in Plants*. Columbia University Press, New
- 4 York, NY, USA.
- 5
- 6 Tenaillon, M. I., J. D. Hollister, and B. S. Gaut. 2010. A triptych of the evolution of plant
- 7 transposable elements. *Trends Plant Sci*. 15:471–478.
- 8
- 9 Trivers, R., A. Burt, and B. G. Palestis. 2004. B chromosomes and genome size in
- 10 flowering plants. *Genome* 47:1–8.
- 11
- 12 Vekemans, X., C. Poux, P. M. Goubet, and V. Castric. 2014. The evolution of selfing
- 13 from outcrossing ancestors in Brassicaceae: what have we learned from variation at the S-
- 14 locus? *J. Evol. Biol.* 27:1372–1385.
- 15
- 16 Wagner, W. L., P. C. Hoch, and P. H. Raven. 2007. Revised classification of the
- 17 Onagraceae. *Syst. Bot. Monogr.* 83:1–222.
- 18
- 19 Whitney, K.D., E. J. Baack, J. L. Hamrick, M. J. W. Godt, B. C. Barringer, M. D.
- 20 Bennett, C. G. Eckert, C. Goodwillie, S. Kalisz, I. J. Leitch, et al. 2010. A role for
- 21 nonadaptive processes in plant genome size evolution? *Evolution* 64:2097–2109.
- 22
- 23 Whitney, K. D., and T. Garland. 2010. Did genetic drift drive increases in genome

- 1 complexity? PLoS Genet. 6.
2
3 Whitney, K. D., B. Boussau, E. J. Baack, and T. Garland. 2011. Drift and genome
4 complexity revisited. PLoS Genet. 7.
5
6 Whitton, J., C. J. Sears, E. J. Baack, and S. P. Otto. 2008. The dynamic nature of
7 apomixis in the angiosperms. Int. J. Plant Sci. 169:169–182.
8
9 Wright, N. A., T. R. Gregory, and C. C. Witt. 2014. Metabolic 'engines' of flight drive
10 genome size reduction in birds. Proc. R. Soc. B 281:20132780.
11
12 Wright, S. I., and D. J. Schoen. 1999. Transposon dynamics and the breeding system.
13 Genetica 107:139–48.
14
15 Wright, S. I., R. W. Ness, J. P. Foxe, and S. C. H. Barrett. 2008. Genomic consequences
16 of outcrossing and selfing in plants. Int. J. Plant Sci. 169:105–118.
17
18 Wright, S. I., and J. A. Ågren. 2011. Sizing up *Arabidopsis* genome evolution. Heredity
19 (Edinb). 107:509–510.
20
21 Zeyl, C., G. Bell, and D. M. Green. 1996. Sex and the spread of retrotransposon *Ty3* in
22 experimental populations of *Saccharomyces cerevisiae*. Genetics 143:1567–77.
23

- 1 Zonneveld, B. J., I. J. Leitch, and M. D. Bennett. 2005. First nuclear DNA amounts in
- 2 more than 300 angiosperms. *Ann. Bot.* 96:229–44.



DNA transposons

LINE

LTR copia

LTR gypsy

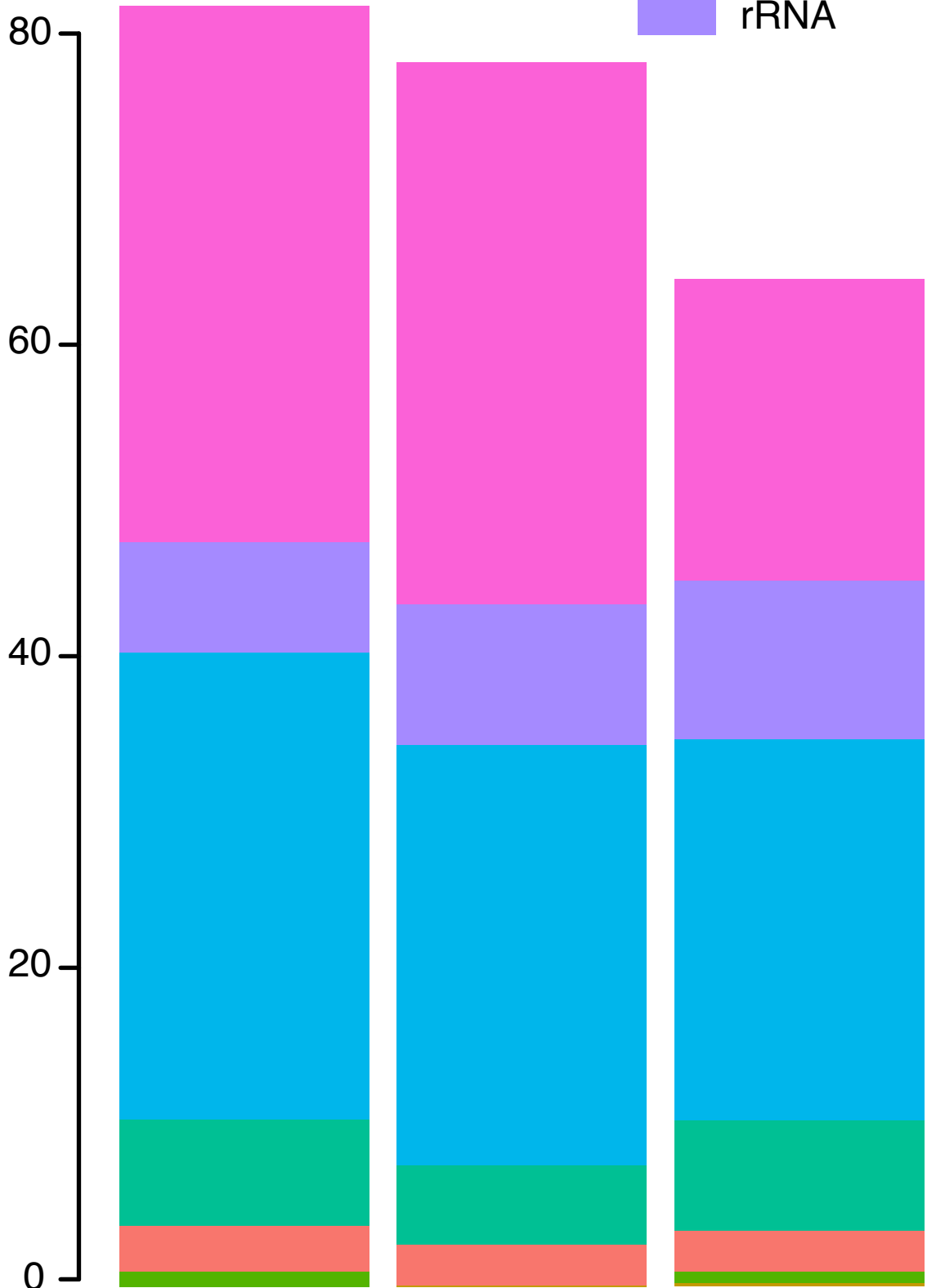
Helitron

LINE

Short simple repeats

rRNA

Estimated Genome Proportion (%)



O. biennis (PTH) *O. elata* (sex) *O. villaricae* (PTH)

