MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 3356–3369 doi: 10.1111/mec.13717

Population genomics of sexual and asexual lineages in fissiparous ribbon worms (Lineus, Nemertea): hybridization, polyploidy and the Meselson effect

S. L. AMENT-VELÁSQUEZ,*† E. FIGUET,* M. BALLENGHIEN,* E. E. ZATTARA,‡§ J. L. NORENBURG,§ F. A. FERNÁNDEZ-ÁLVAREZ,¶ J. BIERNE,** N. BIERNE* and N. GALTIER*

*Institute of Evolutionary Sciences, University Montpellier, CNRS, IRD, EPHE, Place Eugène Bataillon, 34095 Montpellier, France, †Department of Organismal Biology, Systematic Biology, Uppsala University, PO Box 256, SE-751 05 Uppsala, Sweden, ‡Department of Biology, Indiana University, 107 S Indiana Ave, Bloomington, IN 47405, USA, §Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th St. & Constitution Ave. NW, Washington, DC 20560, USA, ¶Institut de Ciències del Mar, CSIC Barcelona, Passeig Marítim, 08003 Barcelona, Spain, **Laboratoire de Biologie Cellulaire et Moléculaire, Université de Reims Champagne-Ardenne, 9 Boulevard de la Paix, 51100 Reims, France*

Abstract

Comparative population genetics in asexual vs. sexual species offers the opportunity to investigate the impact of asexuality on genome evolution. Here, we analyse coding sequence polymorphism and divergence patterns in the fascinating *Lineus* ribbon worms, a group of marine, carnivorous nemerteans with unusual regeneration abilities, and in which asexual reproduction by fissiparity is documented. The population genomics of the fissiparous *L. pseudolacteus* is characterized by an extremely high level of heterozygosity and unexpectedly elevated π_N/π_S ratio, in apparent agreement with theoretical expectations under clonal evolution. Analysis of among-species allele sharing and read-count distribution, however, reveals that *L. pseudolacteus* is a triploid hybrid between Atlantic populations of *L. sanguineus* and *L. lacteus*. We model and quantify the relative impact of hybridity, polyploidy and asexuality on molecular variation patterns in *L. pseudolacteus* and conclude that (i) the peculiarities of *L. pseudolacteus* population genomics result in the first place from hybridization and (ii) the accumulation of new mutations through the Meselson effect is more than compensated by processes of heterozygosity erosion, such as gene conversion or gene copy loss. This study illustrates the complexity of the evolutionary processes associated with asexuality and identifies *L. pseudolacteus* as a promising model to study the first steps of polyploid genome evolution in an asexual context.

Keywords: clonality, coding sequences, hybridization, polyploidy, population genomics

Received 15 December 2015; revision received 18 April 2016; accepted 2 May 2016

Introduction

The vast majority of eukaryotic species reproduce sexually, obligate asexuality being viewed as an evolutionary dead end (Maynard Smith 1978; Simon *et al.* 2003; Zimmer 2009). There are, however, a number of exceptions, that is lineages in which asexual reproduction is the rule (Judson & Normark 1996; Mark Welch & Meselson 2000). Understanding the genomic causes and consequences of the transitions to asexuality is a topic of primary interest, potentially relevant to the long-standing issue of the evolution of sex (Neiman *et al.* 2014).

Population genetics theory provides a number of predictions regarding genome evolution in an asexual context (Birky 1996; Glemin & Galtier 2012; Hartfield 2016). First, in the absence of segregation and genetic exchange between homologous gene copies, individual

Correspondence: Nicolas Galtier,
E-mail: nicolas.galtier@univ-montp2.fr

heterozygosity is expected to increase by accumulation of new mutations – the so-called Meselson effect. As a consequence, increased amounts of neutral genetic polymorphism are expected in clonal lineages, as well as an elevated value of the F_{IS} index, which measures the proportion of within-individual vs. between-individual heterozygosity (Balloux *et al.* 2003). Second, the lack of recombination in asexual lineages is expected to decrease the efficiency of natural selection as a result of genetic hitchhiking, that is interference between selected loci (Gordo & Charlesworth 2001). A selective sweep in an asexual species, for instance, proceeds through the spread of a high-fitness clone, thus increasing the population frequency of not only the favourable, but also the deleterious mutations it carries. Consequently, the ratio of nonsynonymous (selected) to synonymous (neutral) polymorphisms is expected to be higher in asexual than in sexual lineages, because of less efficient purifying selection in the former. Third, because mutations typically occur in heterozygous state in asexual species, recessive mutations are not exposed to selection, which is another reason why a higher load of deleterious mutations is expected in asexual populations.

These theoretical predictions, to date, have only met equivocal empirical validation. The Meselson effect was first described in bdelloid rotifers (Mark Welch & Meselson 2000), but subsequently refuted as new data were generated (Mark Welch *et al.* 2008). In other taxa, the Meselson effect was sometimes observed but could often be explained by mechanisms other than an increase of heterozygosity after the transition to asexuality (e.g. Delmotte *et al.* 2003; Schaefer *et al.* 2006; Lunt 2008; Schwander *et al.* 2011; see Hartfield 2016 for a review). For the efficacy of natural selection, inconsistent results have been reported among groups. A strong effect of asexuality on the nonsynonymous/synonymous substitution rate ratio was reported in various groups of plants, snails and insects (Johnson & Howard 2007; Neiman *et al.* 2010; Henry *et al.* 2012; Hollister *et al.* 2015), but no significant difference was detected between sexuals and asexuals in some aphids (Normark & Moran 2000) and bdelloid rotifers (Mark Welch & Meselson 2001; Barraclough *et al.* 2007). In *Daphnia pulex*, Tucker *et al.* (2013) showed that the deleterious mutations detected in asexuals by Paland & Lynch (2006) did not start to accumulate after the transition to asexuality but were in fact segregating in sexual populations prior to the evolution of asexuality.

So why is there only mixed evidence for a process expected to deeply impact molecular evolution genome-wide? Several explanations can be proposed. First, many of the model asexual species are either very ancient, such as bdelloid rotifers (Mark Welch & Meselson 2000) and darwinulid ostracods (Martens & Schon 2008), and thus lack sexual relatives for comparative purposes, or very recent, such as water fleas *Daphnia pulex* (Tucker *et al.* 2013) and pea aphids *Acyrthosiphon pisum* (Jaquiéry *et al.* 2014), in which genomes are not yet substantially affected by asexual evolution. In a number of supposedly asexual species, evidence for meiosis and recombination was eventually uncovered (Schurko *et al.* 2009; den Bakker *et al.* 2010). Second, most of the data sets that have been analysed in a comparative way so far were based on just a few genes, and/or lacked polymorphism data, and therefore had only limited power.

Most importantly, asexuality is often associated with other genetic peculiarities, including hybridization (Simon *et al.* 2003; Moon *et al.* 2004) and polyploidy (Neiman *et al.* 2014). These are confounding factors potentially contributing to an elevation of heterozygosity, and a modification of selective pressures genome-wide. Of note, the lack of meiotic recombination in asexuals can be alleviated to some extent by gene conversion between alleles, which restores homozygosity (Flot *et al.* 2013), and this process is expected to be particularly prevalent in polyploids (Sémon & Wolfe 2007). The complete genomes of the asexual nematode *Meloidogyne incognita* (Abad *et al.* 2008) and rotifer *Adineta vaga* (Flot *et al.* 2013) have demonstrated the potential of asexual genomes to harbour gene copies more divergent than in typical sexual species, but have not fully clarified the underlying evolutionary processes (Castagnone-Sereno & Danchin 2014). A thorough sampling of genes and species appears to be required to disentangle these complex effects (Hollister *et al.* 2015).

Ribbon worms from the genus *Lineus* and allies represent an interesting opportunity to study the effects of asexuality on genome evolution. This understudied taxon belongs to the widespread phylum Nemertea, which includes over 1200 species of mainly marine worms, which are either carnivores or scavengers (Kajihara *et al.* 2008). *Lineus* worms show remarkable life history and histological traits. *L. longissimus*, for instance, has been reported to reach up to 60 m in length, making it the longest known metazoan (Gittenberger & Schipper 2008). Its maximal growth rate in laboratory conditions is 1 cm per month (J. Bierne, unpublished results), suggesting an extremely long maximal lifespan in this species. Other *Lineus* species, particularly *L. ruber* and *L. sanguineus*, are capable of extensive body regeneration, such that various kinds of chimera have been produced by grafting body parts from distinct individuals, sometimes from distinct species (Bierne 1990).

Interestingly, individuals from two species, *L. sanguineus* and *L. pseudolacteus*, can reproduce by fissiparity: they undergo not only posterior but also anterior regeneration when sectioned in the laboratory (Coe

1929). Fissiparous individuals typically lack mature female gonads, both in the wild and in captivity (Gontcharoff 1951; Bierne 1970). For these reasons, they are suspected to reproduce mainly, if not exclusively, in an asexual way in nature, even though reproduction in the wild has never been observed in *Lineus*. *L. sanguineus* has a worldwide distribution, and several subspecies have been distinguished based on colour pattern and geographic range (Bierne *et al.* 1993), suggesting a relatively ancient origin of asexuality in this cosmopolitan species. *L. pseudolacteus*, which so far has only been found in Brittany (France), still lacks any report of an individual with trace of ovary or testis. Its regenerative abilities are weaker than those of *L. sanguineus*. Both are related to the sexual *L. lacteus* and *L. longissimus* based on morphology, anatomy and immunotransplantations (Bierne *et al.* 1993). This group therefore appears to be an ideal model for the study of genome evolution in asexual *vs.* sexual context – with the additional peculiarity that asexual reproduction in *Lineus* is agametic, in contrast to the other groups of asexuals studied so far. That *L. sanguineus* and *L. pseudolacteus* populations evolve in an asexual context, however, is only suspected based on phenotypic observations, and remains to be confirmed genetically.

The available molecular data in *Lineus* are essentially limited to ribosomal and mitochondrial sequences, which have been generated for large-scale phylogenetic purposes (Struck & Fisse 2008; Andrade *et al.* 2012; Kvist *et al.* 2014). These and previous studies have revealed that current taxonomy is misleading: the genus *Lineus* is not monophyletic, and synonyms are common. The group studied here is monophyletic and includes seven named species – the asexual *L. sanguineus* and *L. pseudolacteus*, and the sexual *L. lacteus*, *L. longissimus*, *Ri. occultus*, *L. viridis* and *L. ruber*. These species have been affiliated to genera *Lineus*, *Myoisophagos*, *Ramphogordius* and *Riseriellus*, but inconsistently so across publications (e.g. Riser 1994; Sundberg & Saur 1998; Kajihara *et al.* 2008; Herrera-Bachiller *et al.* 2015). Despite recent efforts (Kang *et al.* 2015), their taxonomic status is currently unclear, mainly because of insufficient phylogenetic resolution.

Here, we characterized the whole-body transcriptome of 38 individuals from these seven species and sequenced two mitochondrial fragments from 44 additional individuals. Our objectives were to (i) clarify species delineation and taxonomy in this group of nemerteans, (ii) test the hypothesis of predominantly asexual reproduction in the fissiparous *L. sanguineus* and *L. pseudolacteus*, (iii) uncover the origins and age of the asexual lineages and (iv) characterize population genomic processes in asexuals vs. sexuals, with a focus on the Meselson effect, the efficiency of purifying selection, hybridization and polyploidy.

Material and methods

Sampling and sequencing

Eighty two individuals were sampled in the wild in various European and American localities (Table S1, Supporting information). DNA was extracted from the whole body using standard protocols. A fragment of each of the mitochondrial genes cytochrome oxidase 1 (*cox1*) and 16S ribosomal RNA (*16S*) was amplified by PCR (primer sequences for *cox1*: LvLCO1490: 5′-ATTCTACTAATCA TAAAGATATTGG; HCO2198: 5′-TAAACCTCAGGGT GACCAAAAAATCA; *16S*: Lv16sL: 5′-GCTCAACTGTT-TATCAAAAACAT; Lv16sH 5′-GCCGGTCTGAACT CAACTCACGT) and Sanger-sequenced. RNA was extracted from 38 of these samples (whole body) using standard protocols as described by Gayral *et al.* (2011), and non-normalized cDNA libraries were prepared. The libraries were sequenced on HiSeq 2000 (Illumina, Inc.) to produce 100-bp single-end fragments (Table 1). Nine of the transcriptome data sets were previously published as part of a comparative analysis across Metazoa (Romiguier *et al.* 2014). Transcriptome Illumina reads from the outgroup nemertean *Cerebratulus marginatus* (Andrade *et al.* 2014) were downloaded from the SRA database (SRX205323).

Transcriptome assembly

In each species, reads from distinct individuals were gathered and *de novo* transcriptome assembly was performed following strategy B in Cahais *et al.* (2012), yielding one set of predicted cDNA per species. This is a three-step procedure involving one run of ABYSS (Simpson *et al.* 2009) with default parameters, followed by two runs of CAP3 (Huang & Madan 1999) with default parameters, singletons being discarded after step 1. For each contig, open reading frames (ORFs) were predicted using the program GETORF, and the longest ORF was selected. ORFs shorter than 200 bp were discarded. The CD-HIT-EST program (Li *et al.* 2001, default parameters) was applied to the predicted ORFs to assess putative allele splitting – that is several alleles from the same gene assembled as distinct contigs.

Orthology prediction

Orthology between ORFs was predicted by applying the ORTHOMCL pipeline (Li *et al.* 2003) to amino acid-translated sequences (BLAST e-value: 10^{-5}; inflation value: 1.5). Clusters containing exactly one sequence per species were retained. Orthology prediction was performed twice: first using all species (for phylogenomic analyses), and secondly using a subset of six

Table 1 Sampled species and individuals

Lineage	Binomen	Repro.	Sample*	Mreads	ORFs
viridis	*Lineus viridis*	Sexual	1/6	81.5	37 145
ruber_1	*Lineus ruber*	Sexual	1/2	27.3	25 342
ruber_2	*Lineus ruber*	Sexual	2/14	63.3	39 567
occultus	*Riseriellus occultus*	Sexual	1/2	26.8	31 615
longissimus	*Lineus longissimus*	Sexual	6/12	95.6	29 988
lacteus_M	*Lineus (Ramphogordius) lacteus*	Sexual	3/8	95.6	32 307
lacteus_A	*Lineus (Ramphogordius) lacteus*	Sexual	5/17	234.8	67 152
sanguineus	*Lineus (Ramphogordius) sanguineus*	Asexual	16/23	673.1	86 330
pseudolacteus	*Lineus (Ramphogordius) pseudolacteus*	Asexual	3/6	182.7	47 477

*Sample size for transcriptome analysis/mtDNA analysis.

species from the *longissimus* group (for population genetic analyses, see Results).

Phylogenomic analysis

Coding sequences were aligned using the frameshift-aware MACSE program (Ranwez *et al.* 2011) and concatenated. Positions containing >50% of gaps or missing data were removed. Phylogenetic trees were reconstructed using model GTR+Gamma in RAXML (Stamatakis 2014) and performing 1000 bootstrap pseudoreplications. The across-genes average ratio of nonsynonymous to synonymous substitution rate (d_N/d_S) was calculated in each branch of the tree using substitution mapping with the MAPNH program (Romiguier *et al.* 2012).

Genotyping and statistical analyses

Reads were mapped to the whole set of predicted contigs using BWA. Diploid genotypes and single nucleotide polymorphisms (SNPs) were called using the method developed by Tsagkogeorga *et al.* (2012) and implemented in the READS2SNP program. At least 20 reads per individual per position were required to call a genotype; otherwise, missing data were recorded. Paralog filtering in READS2SNP (Gayral *et al.* 2013) was not activated as this method relies on the assumption of Hardy–Weinberg equilibrium, which is clearly not met in asexuals. In each species, only sites at which at least 35% of individuals were successfully genotyped were considered. The average synonymous (π_S) and nonsynonymous (π_N) levels of diversity and the F_{IS} statistics were estimated as in Romiguier *et al.* (2014), discarding sites with more than two alleles. Tajima's D was calculated only based on synonymous sites. Confidence intervals were obtained by bootstrapping loci.

Homologous positions across species were identified based on coding sequence alignments and variants were classified as private polymorphisms (i.e. allelic variation restricted to a single species), shared polymorphisms (i.e. several species carrying the same two alleles) or fixed positions (i.e. one species carrying a specific allele in monomorphic state). Genes showing a strong excess of fixed positions (>10%) in any particular species were discarded as potential remnant paralogs. Subsequent analyses of genotype frequencies within and across species were achieved using home-made python, R and C++ programs relying on the BIO++ libraries (Guéguen *et al.* 2013).

mtDNA analysis

Cox1 and 16S sequences were extracted from each of the 38 transcriptomes and added to the PCR-based mtDNA data set. This was achieved by first making a BLAST search to identify, in each species-level transcriptome assembly, the two contigs carrying the *cox1* and *16S* genes. Then, separately for each individual, reads were mapped to the target contigs with BWA, and a consensus sequence was built to create individual-level haplotypes. The *cox1* and *16S* sequences of outgroup nemertean species *Cerebratulus marginatus*, *Riserius pugetensis* and *Micrura lactea* were included. The former two were retrieved from the NCBI database, while the latter was newly sequenced here. Haplotypes were aligned using the online server of MAFFT (Katoh & Toh 2008; http://mafft.cbrc.jp/alignment/server/). The alignments of *cox1* and *16S* were concatenated, and a phylogenetic tree was built using RAXML, allowing each gene its own model of nucleotide substitution (GTR + G), and performing 1000 bootstrap pseudoreplications.

Results

Cryptic species in Lineus

The mitochondrial tree we obtained (Fig. S1, Supporting information) was consistent with the existing literature

in separating *L. ruber* and *L. viridis* from the *longissimus* group, which includes the sexual *Ri. occultus*, *L. longissimus* and *L. lacteus*, and the asexual *L. sanguineus* and *L. pseudolacteus*. Two highly divergent mitochondrial lineages were found in *L. ruber*, corroborating the existence of two cryptic species (Rogers *et al.* 1995), of which one was more closely related to *L. viridis* than to the other one. In *L. lacteus*, the southern (Mediterranean Sea and Asturias) and northern (Brittany) samples formed two divergent groups of haplotypes, which is suggestive of cryptic speciation in this lineage too. The two *L. ruber* and the two *L. lacteus* lineages were therefore considered as distinct genetic units in the analysis of nuclear data. They were called *ruber*_1, *ruber*_2, *lacteus*_A (Atlantic) and *lacteus*_M (Mediterranean), respectively. For the fissiparous lineages, a single haplotype was obtained in *L. pseudolacteus*, which branched within the homogeneous *L. sanguineus* group, suggesting a close relatedness between these two species.

Transcriptome assembly yielded ~85 000–675 000 contigs per species, of which ~25 000–86 000 contained ORFs longer than 200 bp (Table 1). The number of clusters returned by program CD-HIT-EST was 90.5–99.99% of the number of ORFs, depending on species. This indicates that split alleles are rare. The relatively large numbers of ORFs predicted in some species are best explained by the occurrence of fragmented cDNA. Orthology search across the whole set of species identified 604 groups of predicted orthologues, each containing exactly one sequence per species. The resulting phylogenetic tree (Fig. 1), which was based on 524 094 aligned sites, corroborated the mtDNA analysis and provided 100% bootstrap support for every internal node. The mtDNA and nuclear trees did not support splitting the group into the genera *Ramphogordius* (*R. sanguineus* and *R. pseudolacteus*), *Riseriellus* (*Ri. occultus*) and *Lineus*, as the latter would be paraphyletic. The average nuclear coding sequence divergence between the two *L. ruber* and between the two *L. lacteus* entities were elevated (5% and 3%, respectively), in line with the hypothesis of cryptic species. The d_N/d_S ratio was around 0.13 and showed little variation among branches (Fig. 1). In particular, no increase in d_N/d_S ratio was detected in the branches leading to the fissiparous lineages.

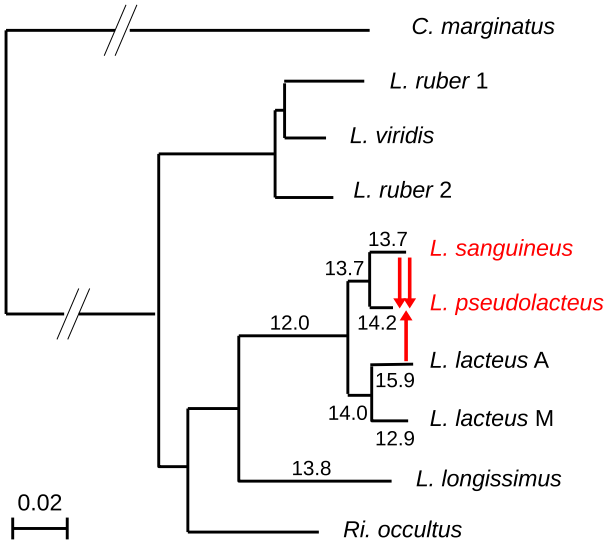

Fig. 1 *Lineus* phylogenomics. Tree topology and branch lengths were obtained by maximum-likelihood analysis of 604 nuclear protein-coding genes. Branch-specific labels correspond to estimates of the d_N/d_S ratio obtained in the *longissimus* group based on a larger set of 2895 genes, multiplied by 100. Fissiparous taxa are in red. Arrows represent the inferred hybridization event from which triploid *L. pseudolacteus* originated.

Read counts and ploidy level

For each individual of each species, reads were mapped to species-level contigs using BWA. For each position at which (i) at least 30 reads were available and (ii) at least two distinct bases had been called, the frequency of the second-most common base was computed. In 33 of 38 individuals, a distribution similar to Fig. 2a was obtained, with a mode at very low frequency presumably reflecting sequencing errors, and a mode around 0.5 corresponding to allelic variation, as expected in a diploid individual. Five samples from candidate asexual lineages provided very different patterns. In each of the three *L. pseudolacteus* individuals (Fig. 2b), one mode was close to 0.3, not 0.5, which is suggestive of polyploidy. Such a distribution seems compatible with triploidy, in which case minor allelic variants at within-individual frequency 1/3 are expected, or tetraploidy, that is a mixture of variants at frequency ¼ and ½. Two *L. sanguineus* individuals – sample Ore_Edu from Oregon and Por_3 from Brittany – showed a very unexpected read-count distribution (Fig. 2c), which might be suggestive of high levels of ploidy and a wide range of within-individual allele frequencies.

To investigate further the hypothesis of polyploidy in *L. pseudolacteus* and two *L. sanguineus* individuals, we looked for the presence of SNPs with three or four alleles in the same individual. We found that the number of positions in which the third most common allele was represented by more than five reads was below five in all individuals, excepting the three *L. pseudolacteus* and the Ore_Edu and Por_3 *L. sanguineus*, in which 18–299 triallelic SNPs were found, corroborating the hypothesis of polyploidy for these five samples. The variation in number of triallelic SNPs between the putative polyploid individuals is explained by coverage: the

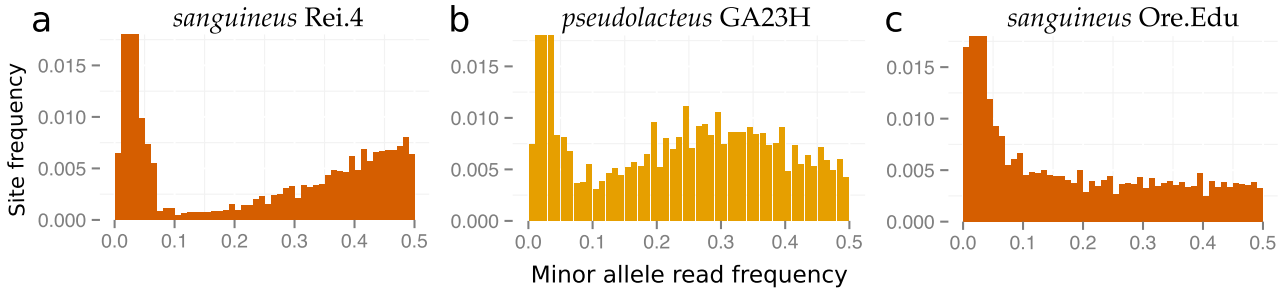

Fig. 2 Minor base read-count distribution. Typical across-sites distributions of the frequency of the second-most common base obtained by mapping to contigs the short read sequences from single individuals.

percentage of triallelic sites among sites with coverage above 30X was quite homogeneous – from 0.023% to 0.047% – across the five individuals. Three tetra-allelic sites were found in the *L. sanguineus* sample from Oregon (Ore_Edu).

Population genomics of sexual vs. asexual lineages

Next, we focussed on the *longissimus* group, which contains the two fissiparous lineages and their sexual relatives. Orthology search within this group yielded 2895 genes for which exactly one copy was available in each of *Ri. occultus*, *L. longissimus*, *L. lacteus*_A, *L. lacteus*_M, *L. sanguineus* and *L. pseudolacteus*. The following analyses were applied to this set of orthologues, to the exclusion of other genes. The *Ri. occultus* lineage, represented by a single individual in this data set, was not considered in population genetic analyses. Diploid genotypes were called using the READS2SNP program (Tsagkogeorga *et al.* 2012). This is appropriate for the diploid individuals, and presumably a reasonable approximation for the potentially polyploid individuals, knowing that only a very small number of positions (<0.1%) were obviously not biallelic, that is showed three bases at relatively high read frequency within the same individual – such positions were discarded.

Standard population genomic summary statistics were calculated separately for each species of the *longissimus* group (Table 2). The average level of neutral genetic diversity, here based on synonymous sites (π_S), was above 1% in the six analysed species, which is typical of many invertebrate species (Romiguier *et al.* 2014). It was maximal and reached 10% in the fissiparous *L. pseudolacteus*. This is higher than the maximal π_S value reported in a recent comparative study of 90 animal species (Romiguier *et al.* 2014). The across-contigs distribution of heterozygosity in four of the analysed species is shown in Fig. S2 (Supporting information). The F_{IS} statistics were not significantly different from zero in the three sexual species, revealing no obvious departure from panmixia, but were significantly negative in the two fissiparous lineages. In particular, it was close to −1, its theoretical minimum, in *L. pseudolacteus*. A negative F_{IS} is indicative of a higher level of heterozygosity within than between individuals, as expected in asexual populations (Balloux *et al.* 2003).

We computed in each species the site frequency spectrum, that is the across-SNPs distribution of minor allele frequencies (Fig. S3, Supporting information), and calculated Tajima's D (Table 2). Site frequency spectra in sexual species were similar to published spectra (e.g. see Gayral *et al.* 2013) and yielded a negative Tajima's D – that is an excess of low-frequency variants compared to an ideal Wright–Fisher population. The *L. pseudolacteus* distribution of allele frequency, in contrast, was highly unusual: the vast majority of SNPs were such that every sampled individual was heterozygous, meaning that allele frequency was 0.5 (Fig. S3,

Table 2 Population genomic analysis, species from the *longissimus* group

Species	#ind	π_S	π_N/π_S	F_{IS}	Tajima's D
L. longissimus	6	0.018 [±0.001]	0.073 [±0.008]	−0.017 [±0.02]	−0.52
*L. lacteus*_M	3	0.015 [±0.001]	0.092 [±0.012]	$6.6\ 10^{-5}$ [±0.03]	−0.25
*L. lacteus*_A	5	0.037 [±0.001]	0.053 [±0.004]	−0.007 [±0.01]	−0.65
L. sanguineus	16	0.032 [±0.001]	0.054 [±0.004]	−0.10 [±0.01]	0.27
L. pseudolacteus	3	0.101 [±0.003]	0.071 [±0.004]	−0.92 [±0.01]	2.26

Table 3 Among-species polymorphism sharing

	L. longissimus	*L. lacteus*_M	*L. lacteus*_A	*L. sanguineus*	*L. pseudolacteus*
Fixed	3540	213	24	24	2
Private	794	386	1014	775	94
Shared with					
L. longissimus		12	29	31	3
*L. lacteus*_M			37	16	3
*L. lacteus*_A				49	187
L. sanguineus					557

Supporting information). The *L. sanguineus* pattern was not nearly as extreme as the *L. pseudolacteus* one, but still revealed a slight excess of high-frequency variants.

Hybrid origin of Lineus pseudolacteus

To further clarify the relationships between the distinct gene pools, we analysed patterns of polymorphism sharing among species, that is sites at which the same two alleles are observed in more than one species. In this analysis, we only considered sites successfully genotyped in all six species of the *longissimus* group (>35% of individuals per species). Private polymorphisms and fixed positions were also counted (Table 3). Only limited numbers of shared polymorphism were found between *L. longissimus*, *L. lacteus*_A, *L. lacteus*_M and *L. sanguineus*, confirming the hypothesis of limited gene flow between the Atlantic and Mediterranean populations of *L. lacteus*. In *L. pseudolacteus*, in contrast, the vast majority of polymorphisms were shared with either *L. sanguineus* or *L. lacteus*_A, with private polymorphisms and, particularly, fixed positions being scarce (Table 3). This result suggests that the polyploid *L. pseudolacteus* taxon results from hybridization between *L. sanguineus* and *L. lacteus*_A. The principal component analysis of the genotype table was consistent with this suggestion (Fig. S4, Supporting information). Individuals clustered by species in this analysis, with *L. pseudolacteus* individuals being located between *L. sanguineus* and *L. lacteus_A* individuals. Principal component analysis revealed no geographic clustering within *L. sanguineus* (Fig. S4c, Supporting information).

To investigate more deeply the hypothesis of a hybrid origin, we specifically considered, for each biallelic polymorphism in *L. pseudolacteus*, the diversity at the same position in *L. sanguineus* and *L. lacteus*_A. In Table 4, X and Y stand for the two alleles segregating in *L. pseudolacteus*, and categories of *L. pseudolacteus* SNPs were defined based on the presence or absence of these alleles in the two putative parental species. Here, we analysed sites successfully genotyped in each of the three species of interest, irrespective of their availability in other *Lineus* species, thus increasing the number of SNPs compared to the analyses above.

Table 4 Polymorphisms in *L. pseudolacteus* and its parental species

Alleles in *L. pseudolacteus*	Alleles in *L. sanguineus*	Alleles in *L. lacteus*_A	SNP number	SNP category
X,Y	X	X	358	Private: truly private
X,Y	X	Y	1631	Private: hybrid
X,Y	X,Y	X	1944	Shared
X,Y	X	X,Y	802	Shared
X,Y	X,Y	X,Y	295	Shared
X,Y	Otherwise		273	Other

Interestingly, besides shared polymorphisms, we found that a large number (1631) of *L. pseudolacteus* private SNPs involved two alleles that were differentially fixed in *L. sanguineus* and *L. lacteus*_A (Table 4) – these were called 'hybrid SNPs'. The elevated proportion of such positions, which form the majority of private SNPs, strongly argues in favour of the hypothesis of a *L. sanguineus* × *L. lacteus*_A hybrid origin for *L. pseudolacteus*. Focussing on the above-defined hybrid SNPs, we examined read counts and calculated the proportion of reads in *L. pseudolacteus* corresponding to the *L. lacteus*_A allele (Fig. 3). We found that the contribution of *L. lacteus*_A alleles to *L. pseudolacteus* reads at hybrid SNPs was unimodal with a median of 0.325, that is very close to 1/3, strongly suggesting that *L. pseudolacteus* is a triploid hybrid made of one haplome of *L. lacteus*_A origin and two haplomes of *L. sanguineus* origin.

Meselson effect in Lineus pseudolacteus

The vast majority of *L. pseudolacteus* SNPs were either hybrid SNPs or SNPs shared with one of its parental

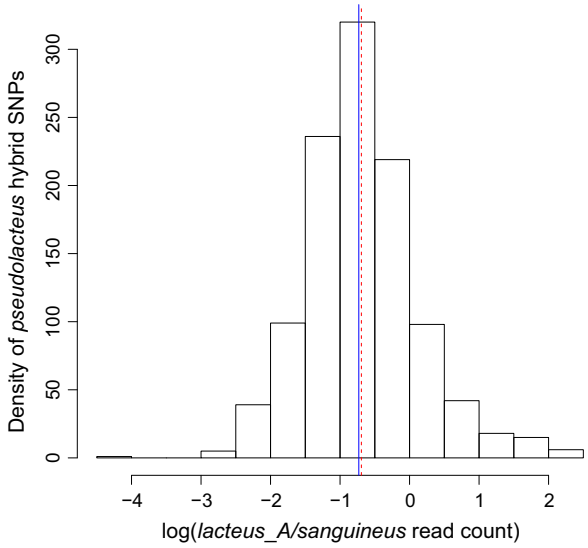

Fig. 3 Relative contributions of *L. lacteus*_A and *L. sanguineus* to *L. pseudolacteus* polymorphism. For each *L. pseudolacteus* hybrid SNP, the number of reads corresponding to the *L. lacteus*_A vs. *L. sanguineus* alleles was recorded. Here is the across-SNPs distribution of the *L. lacteus*_A/*L. sanguineus* read count ratio. Blue vertical line: median of the distribution; red dotted line: expected *L. lacteus*_A/*L. sanguineus* ratio under triploidy.

species (Table 4). Still, 358 SNPs involved an allele that was uniquely found in *L. pseudolacteus*. These polymorphisms, hereafter called 'truly private SNPs', might correspond to new mutations having arisen in this lineage. However, there is also the possibility that such *L. pseudolacteus*-specific alleles were inherited from one of the parental lineages at the time of hybridization, but subsequently lost, or undetected, in the parental species. We took a theoretical approach to assess the relative likelihood of these processes (see Appendix I).

We estimated that ~40% of the truly private SNPs correspond to polymorphisms that are actually shared with *L. lacteus*_A or *L. sanguineus*, but undetected in these species due to limited sample size. As far as the remaining 60% are concerned, we found that the proportion of truly private polymorphisms resulting from new mutations in *L. pseudolacteus* is expected to be ½ in the absence of gene conversion and gene copy loss, or above ½ if either of these processes applies. This prediction is independent of the effective population size of the parental and hybrid species, but assumes that the mutation rate and effective population sizes remain constant through time. This indicates that a substantial proportion of truly private polymorphisms observed in *L. pseudolacteus* is explained by the accumulation of heterozygosity through new mutations in the *L. pseudolacteus* lineage, which is consistent with the Meselson effect.

Truly private SNPs amount to ~0.04% of synonymous positions in *L. pseudolacteus*, indicating a relatively recent origin of this hybrid lineage. Assuming a clockwise accumulation of new mutations in time and no loss of polymorphism in *L. pseudolacteus*, this figure indicates that the hybridization event was 250–500 times more recent than the date of divergence between *L. sanguineus* and *L. lacteus*, which differ by ~13% at synonymous sites. We lack a calibration of the molecular clock in *Lineus* to convert this figure into absolute time. Assuming a per-year, per-site nuclear mutation rate of 10^{-9}, similarly to human and apes (Yi *et al.* 2002), would result in a hybridization date of ~12 000–25 000 years. A higher mutation rate would imply a more recent date of hybridization, while any process of heterozygosity loss in *L. pseudolacteus*, such as gene conversion or gene copy loss, would imply a more ancient date of hybridization.

Efficacy of purifying selection

The across-genes average ratio of nonsynonymous to synonymous heterozygosity, π_N/π_S, was calculated in each species. The π_N/π_S ratio was negatively related to log-transformed π_S (Fig. 4, left). This is consistent with previous reports (Gayral *et al.* 2013; Romiguier *et al.* 2014) and with the nearly neutral theory (Ohta 1987; Lynch 2007), according to which slightly deleterious nonsynonymous mutations are more likely to increase in frequency in small or perturbed populations, where genetic

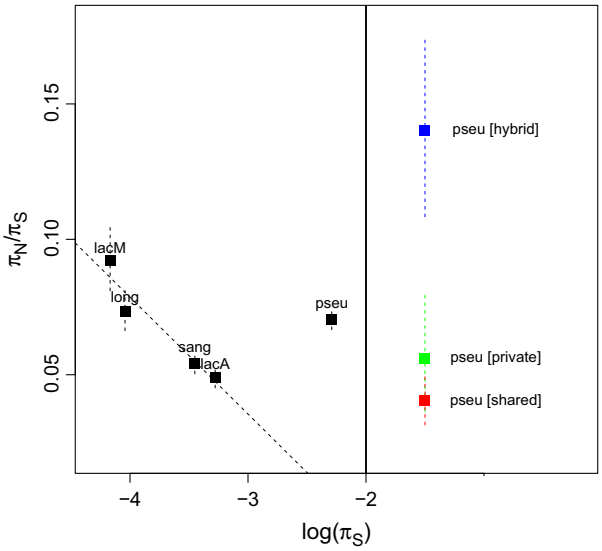

Fig. 4 π_N/π_S ratio in the *longissimus* group. Left part: π_N/π_S is inversely related to log-transformed π_S across species, *L. pseudolacteus* being an outlier. Right part: π_N/π_S varies substantially among categories of *L. pseudolacteus* SNPs. In the right part, the *x*-axis is irrelevant.

drift is stronger. *L. pseudolacteus* appeared to be an outlier of the relationship: the π_N/π_S ratio in *L. pseudolacteus* was higher than expected given its level of neutral diversity.

However, when we calculated π_N/π_S separately for the three categories of SNPs, it appeared that the elevated π_N/π_S in *L. pseudolacteus* is mainly caused by hybrid SNPs (Fig. 4, right). Hybrid SNPs consist of two alleles that were differentially fixed during the divergence between *L. sanguineus* and *L. lacteus*_A. These positions therefore convey information about the long-term process of coding sequence divergence, not polymorphism. The high nonsynonymous/synonymous ratio at hybrid SNPs in *L. pseudolacteus*, therefore, merely reflects the fact that d_N/d_S is generally higher than π_N/π_S in *Lineus* nemerteans (see Fig. 1 and Table 2). The shared and truly private SNPs, in contrast, showed moderate π_N/π_S levels.

Discussion

Species delineation and taxonomy in Lineus

Our mitochondrial and nuclear data corroborated the existence of two cryptic species in *L. ruber*, as suggested by Rogers *et al.* (1995) from 13 allozyme loci. Importantly, we uncovered a cryptic event of speciation within *L. lacteus*. Samples from the Mediterranean Sea and the Asturias clustered separately from the Northern Atlantic samples, and a very limited amount of gene flow was detected between the two groups. Additional sampling is required to characterize further the geographic range of the newly identified gene pools, especially in the putative zone of sympatry in northern Spain.

Phylogenetic analysis consistently supported the grouping of *L. viridis* and *L. ruber* vs. all other *Lineus* species, in agreement with morphological and immunotransplantation data (Bierne *et al.* 1993). The *longissimus* group contained species sometimes assigned to genera *Ramphogordius* (*R. sanguineus*, *R. pseudolacteus*) and *Riseriellus* (*Ri. occultus*), which, if validated, would make the genus *Lineus* paraphyletic (Fig. 1). Given that *L. longissimus* is the type species for the genus *Lineus*, it is reasonable at this time to regard the *Ramphogordius* (=*Myoisophagos*) species as *Lineus*, whereas taxonomic placement of *Ri. occultus, L. ruber* and *L. viridis* remains moot and beyond the scope of this study.

Sexual-like population genomics in the fissiparous Lineus sanguineus

Lineus sanguineus individuals present remarkable regeneration abilities, including the ability to regenerate the head when cut transversally (Coe 1929). This property and the extreme rarity of mature females, both in the wild and in the laboratory, have led to the suggestion that *L. sanguineus* reproduces only by fissiparity, and is therefore an asexual lineage (Gontcharoff 1951; Riser 1994). Our population genomic analysis failed to firmly confirm this hypothesis. The sixteen *L. sanguineus* individuals we sequenced were genetically distinct from each other, revealed no geographic structure, and harboured within-individual levels of heterozygosity comparable to sexual *L. lacteus* individuals. The negative F_{IS} and positive Tajima's D statistics are both suggestive of a departure from panmixia and polymorphism retention in *L. sanguineus*, possibly because of asexuality, but the evidence is not strong. Importantly, the theory predicts that small amounts of sexual reproduction are sufficient to essentially erase the expected population genetic signal associated with asexuality (Balloux *et al.* 2003; Hartfield *et al.* 2016). It seems therefore plausible that fissiparity is the dominant mode of reproduction in *L. sanguineus*, but that sexual reproduction occurs in this species at a sufficient rate to promote genetic exchange between individuals and across loci. A deeper sampling strategy would be required to estimate the prevalence of asexual vs. sexual reproduction in current and past populations of *L. sanguineus*.

Origin and evolution of the asexual Lineus pseudolacteus

The distribution of heterozygosity across *L. pseudolacteus* individuals, in contrast, was fully consistent with the hypothesis of clonal evolution: the vast majority of polymorphisms were shared among the three sampled individuals, so that F_{IS} was close to -1 in this lineage, in which not a single individual with mature gonads has been observed so far. It is therefore likely that natural populations of *L. pseudolacteus* only experience asexual reproduction, offering an opportunity to investigate molecular evolutionary patterns in the absence of sex.

At first sight, the population genomics of *L. pseudolacteus* meets all the theoretical predictions associated with clonal evolution. First, the level of genetic polymorphism is extremely high in this species, in agreement with the Meselson effect hypothesis. The genome-average synonymous diversity, π_S, which reaches 10%, is among the highest ever reported in any eukaryotic species (Small *et al.* 2007; Dey *et al.* 2013; Romiguier *et al.* 2014). Second, the π_N/π_S ratio is higher in *L. pseudolacteus* than in its sexual relatives *L. sanguineus* and *L. lacteus*_A, in agreement with the hypothesis of a reduced efficiency of natural selection in asexual lineages.

A closer examination of the data, however, reveals that *L. pseudolacteus* is very likely to be of hybrid origin. Patterns of polymorphism sharing and read counts

strongly suggest that *L. pseudolacteus* individuals are triploid, their genome being made of two haplomes inherited from a *L. sanguineus* ancestor and one haplome inherited from a *L. lacteus*_A ancestor. The mtDNA tree, in which *L. pseudolacteus* is nested within *L. sanguineus*, indicates that *L. pseudolacteus* inherited the *L. sanguineus* mitochondrial genome. *L. pseudolacteus* might therefore have originated through the fertilization by a *L. lacteus*_A male gamete of an unreduced *L. sanguineus* female gamete. Both *L. sanguineus* and *L. lacteus*_A occur along the eastern Atlantic coasts, and especially in Brittany, where current *L. pseudolacteus* populations are located. An accurate estimation of the hybridization date is out of reach in the absence of fossil calibration, but conservative assumptions regarding plausible mutation rates in *Lineus* indicate that the hybridization event probably took place during the last 100 000 years. Hybridity, polyploidy and asexuality are commonly found in association in nature (Mogie 1986; Lampert & Schartl 2008; Lovell *et al.* 2013), a typical interpretation being that allopolyploidy might hamper proper meiosis and force asexuality.

The impact of asexuality on *L. pseudolacteus* molecular evolution should therefore be appreciated in the light of its hybrid origin. Regarding the level of genetic diversity, we found that most of *L. pseudolacteus* heterozygosity was either inherited from one of its parental species or created by the admixture of the two diverged genomes. So the high level of genetic polymorphism in *L. pseudolacteus* is explained in the first place by its hybrid nature, not asexuality. Still, we identified a portion of truly private SNPs that correspond to new mutations in the *L. pseudolacteus* lineage, substantiating the Meselson effect. This accretion of new mutations has not yet deeply impacted the *L. pseudolacteus* genome, asexuality being too recent.

On the other hand, we note that π_S in *L. pseudolacteus* (10%) is lower than the synonymous divergence between its two parental species (13%), which is an estimate of synonymous heterozygosity in the ancestral hybrid. This indicates that part of the heterozygosity initially created when the hybrid was formed has been lost during *L. pseudolacteus* evolution. Gene conversion and gene copy loss are typical mechanisms of heterozygosity erosion in asexuals (Flot *et al.* 2013). Because we are here analysing transcriptome-based data, allele silencing is another hypothesis to be considered (Pala *et al.* 2010). The apparent homozygosity has increased in *L. pseudolacteus* at a much faster rate (~3% of heterozygosity lost since the ancestral hybrid) than the accumulation of new mutations through the Meselson effect (0.012–0.025%), implying either a rate of gene conversion/gene copy loss/silencing ~100 times as high as the point mutation rate or, perhaps more plausibly, adaptive homozygosity recovery – natural selection is expected to favour the loss of heterozygosity at loci involved in genetic incompatibility between the two parental genomes.

The confounding effect of hybridization

Regarding the π_N/π_S ratio, distinguishing between categories of SNPs revealed that the relative high average π_N/π_S in *L. pseudolacteus* is entirely caused by hybrid SNPs, that is alleles differentially fixed in *L. sanguineus* vs. *L. lacteus*_A. The genetic variation at hybrid SNPs has been established in a sexual context, as the two parental species were diverging. For this reason, the ratio of nonsynonymous to synonymous changes at these positions reflects the long-term influence of both purifying and positive selection, and is similar to the d_N/d_S estimates we report between *Lineus* species, as expected. d_N/d_S is higher than π_N/π_S in *Lineus*, as previously reported in invertebrate species with large population sizes (Smith & Eyre-Walker 2002; Tsagkogeorga *et al.* 2012). The elevated π_N/π_S in *L. pseudolacteus*, therefore, should not be taken as evidence for a relaxed efficiency of purifying selection in this lineage. A similar pattern was recently reported in the plant genus *Oenothera* (Hollister *et al.* 2015), in which π_N/π_S ratio in hybrid diploid asexuals was similar to the d_N/d_S ratio measured in sexual lineages (their fig. 2).

Besides hybrid SNPs, shared and truly private SNPs in *L. pseudolacteus* did not reveal any increase of the π_N/π_S ratio compared to its sexual relatives, failing to corroborate the prediction of relaxed purifying selection in asexuals. This is perhaps not so unexpected given the relatively young age of the *L. pseudolacteus* lineage. Under clonal evolution, shared SNPs, which were present in the ancestral hybrid, are expected to be carried at polymorphic state by all *L. pseudolacteus* individuals. No increase in π_N/π_S ratio is to be expected for these sites – they do not contribute any fitness difference between individuals. The argument also applies to the fraction of truly private SNPs – approximately one half – that corresponds to ancestral heterozygosity subsequently lost in the parental species. An increase in π_N/π_S ratio due to less efficient purifying selection is therefore only expected in one half of the truly private SNPs – a tiny fraction of *L. pseudolacteus* SNPs. Figure 4 suggests that such an increase might well have occurred, the π_N/π_S ratio being higher at truly private than at shared SNPs, but the power of the analysis is insufficient for a firm conclusion on this issue – especially knowing that a substantial fraction of our 'truly private' SNPs actually correspond to missed shared polymorphisms.

Concluding remarks

A significant increase in heterozygosity and π_N/π_S ratio was detected in the asexual *L. pseudolacteus*, compared to its sexual relatives, in agreement with theoretical expectations. We show, however, that these peculiarities are explained in the first place by the hybrid and polyploid nature of *L. pseudolacteus*. Asexuality has indeed affected the population genetics of *L. pseudolacteus*, but the genomewide impact is limited owing to its relatively recent origin. Differentiating between hybridization and drift effects was only possible thanks to our exhaustive taxon sampling, multilocus molecular coverage, and focussed data analysis. We suggest that such an in-depth characterization of population genomic patterns might generally be required to discriminate the impact of asexuality from that of confounding factors, particularly hybridization, as previously discussed in other asexual systems (e.g. Castagnone-Sereno & Danchin 2014).

This study identifies *L. pseudolacteus* as a promising biological model to study the first steps of adaptive and nonadaptive genome evolution in an asexual context. The *Lineus* system is original in involving polyploidy and agametic reproduction. Polyploidy is expected to make a difference in terms of gene conversion/allele loss/allele silencing rate, every gene being present in multiple copies (Renny-Byfield & Wendel 2014). Agametic reproduction implies that 'somatic' mutations can be transmitted from generation to generation and yield mosaic organisms, with important consequences with respect to ageing (Nilsson Sköld & Obst 2011). The consequences of this peculiar genetic architecture on genome and mating system evolution are worth investigating deeper.

Acknowledgements

This work was supported by European Research Council grant 232971 (PopPhyl), Agence Nationale de la Recherche grants ANR-10-BINF-01-01 (Ancestrome) and ANR-11-BSV7-007 (Clonix), the Erasmus Mundus Master Programme in Evolutionary Biology, a University of Maryland and Smithsonian Institution Seed Grant for Research, and Ministry of Economy and Competitiveness of Spain predoctoral grant BES-2013-063551. We thank the Montpellier Bioinformatics & Biodiversity platform and the ISEM platform PGPM7 at the Station marine (OSU OREME) for support. We thank Laurent Léveque, Delphine Lallias, Marguerite Bierne, Benoit Houseaux and Christophe Lemaire for their help with sampling. We are grateful to Sophie Arnaud-Haond and four reviewers for useful suggestions.

References

Abad P, Gouzy J, Aury J-M *et al.* (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nature Biotechnology*, **26**, 909–915.

Andrade SCS, Strand M, Schwartz M *et al.* (2012) Disentangling ribbon worm relationships: multi-locus analysis supports traditional classification of the phylum Nemertea. *Cladistics*, **28**, 141–159.

Andrade SC, Montenegro H, Strand M *et al.* (2014) A transcriptomic approach to ribbon worm systematics (Nemertea): resolving the Pilidiophora problem. *Molecular Biology and Evolution*, **31**, 3206–3215.

den Bakker HC, Vankuren NW, Morton JB, Pawlowska TE (2010) Clonality and recombination in the life history of an asexual arbuscular mycorrhizal fungus. *Molecular Biology and Evolution*, **27**, 2474–2486.

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.

Barraclough TG, Fontaneto D, Ricci C, Herniou EA (2007) Evidence for inefficient selection against deleterious mutations in cytochrome oxidase I of asexual bdelloid rotifers. *Molecular Biology and Evolution*, **24**, 1952–1962.

Bierne J (1970) Recherche sur la différenciation sexuelle au cours de l'ontogenèse et de la régénération chez le Némertien Lineus ruber (Muller). *Annales des Sciences Naturelles, Zoologie*, **12**, 181–298.

Bierne J (1990) *Lineus* as a model for studying developmental processes in animals reconstructed from adult pieces. *International Journal of Developmental Biology*, **34**, 245–253.

Bierne J, Tarpin M, Vernet G (1993) A reassessment of the systematics and a proposal for the phylogeny of some cosmopolitan *Lineus* species (Nemertea). *Hydrobiologia*, **266**, 159–168.

Birky CW (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics*, **144**, 427–437.

Cahais V, Gayral P, Tsagkogeorga G *et al.* (2012) Reference-free transcriptome assembly in non-model animals from next generation sequencing data. *Molecular Ecology Resources*, **12**, 834–845.

Castagnone-Sereno P, Danchin E (2014) Parasitic success without sex – the nematode experience. *Journal of Evolutionary Biology*, **27**, 1323–1333.

Coe WR (1929) Regeneration in nemerteans. *Journal of Experimental Zoology*, **54**, 411–459.

Delmotte F, Sabater-Munoz B, Prunier-Leterme N *et al.* (2003) Phylogenetic evidence for hybrid origins of asexual lineages in an aphid species. *Evolution*, **57**, 1291–1303.

Dey A, Chan CK, Thomas CG, Cutter AD (2013) Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 11056–11060.

Flot J-F, Hespeels B, Li X *et al.* (2013) Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature*, **500**, 453–457.

Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in non-model animals. *Molecular Ecology Resources*, **11**, 650–661.

Gayral P, Melo-Ferreira J, Glémin S *et al.* (2013) Reference-free population genomics from Next-Generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics*, **9**, e10003457.

Gittenberger A, Schipper C (2008) Long live Linnaeus, *Lineus longissimus* (Gunnerus, 1770) (Vermes: Nemertea: Anopla:

Heteronemertea: Lineidae), the longest animal worldwide and its relatives occurring in The Netherlands. *Zoologische Mededelingen Leiden*, **82**, 59–63.

Glemin S, Galtier N (2012) Genome evolution in outcrossing versus selfing versus asexual species. *Methods in Molecular Biology*, **855**, 311–335.

Gontcharoff M (1951) Biologie de la régénération chez quelques Linaeidae de France. *Annales des Sciences Naturelles, Zoologie*, **13**, 149–235.

Gordo I, Charlesworth B (2001) Genetic linkage and molecular evolution. *Current Biology*, **11**, R684–R686.

Guéguen L, Gaillard S, Boussau B *et al.* (2013) Bio++: efficient, extensible libraries and tools for molecular evolution. *Molecular Biology and Evolution*, **30**, 1745–1750.

Hartfield M (2016) Evolutionary genetic consequences of facultative sex and outcrossing. *Journal of Evolutionary Biology*, **29**, 5–22.

Hartfield M, Wright SI, Agrawal AF (2016) Coalescent times and patterns of genetic diversity in species with facultative sex: effects of gene conversion, population structure and heterogeneity. *Genetics*, **202**, 297–312.

Henry L, Schwander T, Crespi BJ (2012) Deleterious mutation accumulation in asexual *Timema* stick insects. *Molecular Biology and Evolution*, **29**, 401–408.

Herrera-Bachiller A, Fernández-Álvarez FA, Junoy J. (2015) A Taxonomic Catalogue of the nemerteans (Phylum Nemertea) from Spain and Portugal. *Zoological Science*, **32**, 507–522.

Hollister JD, Greiner S, Wang W *et al.* (2015) Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. *Molecular Biology and Evolution*, **32**, 896–905.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research*, **9**, 868–877.

Jaquiéry J, Stoeckel S, Larose C *et al.* (2014) Genetic control of contagious asexuality in the pea aphid. *PLoS Genetics*, **10**, e1004838.

Johnson SG, Howard RS (2007) Contrasting patterns of synonymous and nonsynonymous sequence evolution in asexual and sexual freshwater snail lineages. *Evolution*, **61**, 2728–2735.

Judson OP, Normark BB (1996) Ancient asexual scandals. *Trends in Ecology & Evolution*, **11**, 41–46.

Kajihara H, Chernyshev AV, Sun S-C, Sundberg P, Crandall FB (2008) Checklist of nemertean genera and species published between 1995 and 2007. *Species Diversity*, **13**, 245–274.

Kang X-X, Fernández-Álvarez FA, Alfaya J *et al.* (2015) Species diversity of *Ramphogordius sanguineus/Lineus ruber* like nemerteans (Nemertea: Heteronemertea) and geographic distribution of *R. sanguineus*. *Zoological Science*, **32**, 579–589.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.

Kvist S, Laumer CE, Junoy J, Giribet G (2014) New insights into the phylogeny, systematics and DNA barcoding of Nemertea. *Invertebrate Systematics*, **28**, 287–308.

Lampert KP, Schartl M (2008) The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **363**, 2901–2909.

Li W, Jaroszewski L, Godzik A (2001) Clustering of highly homologous sequences to reduce the size of large protein database. *Bioinformatics*, **17**, 282–283.

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**, 2178–2189.

Lovell JT, Aliyu OM, Mau M *et al.* (2013) On the origin and evolution of apomixis in *Boechera*. *Plant Reproduction*, **26**, 309–315.

Lunt DH (2008) Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evolutionary Biology*, **8**, 194.

Lynch M (2007) *The Origins of Genome Architecture*. Sinauer Associates Inc., Sunderland, MA.

Mark Welch D, Meselson M (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science*, **288**, 1211–1215.

Mark Welch DB, Meselson MS (2001) Rates of nucleotide substitution in sexual and anciently asexual rotifers. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 6720–6724.

Mark Welch DB, Mark Welch JL, Meselson M (2008) Evidence for degenerate tetraploidy in bdelloid rotifers. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5145–5149.

Martens K, Schon I (2008) Ancient asexuals: darwinulids not exposed. *Nature*, **453**, 587.

Maynard Smith J (1978) *The Evolution of sex*. Cambridge University Press, Cambridge.

Mogie M (1986) On the relationship between asexual reproduction and polyploidy. *Journal of Theoretical Biology*, **122**, 493–498.

Moon CD, Craven KD, Leuchtmann A, Clement SL, Schardl CL (2004) Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. *Molecular Ecology*, **13**, 1455–1467.

Neiman M, Hehman G, Miller JT, Logsdon JM, Taylor DR (2010) Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Molecular Biology and Evolution*, **27**, 954–963.

Neiman M, Sharbel TF, Schwander T (2014) Genetic causes of transitions from sexual reproduction to asexuality in plants and animals. *Journal of Evolutionary Biology*, **27**, 1346–1359.

Nilsson Sköld H, Obst M (2011) Potential for clonal animals in longevity and ageing studies. *Biogerontology*, **12**, 387–396.

Normark BB, Moran NA (2000) Testing for the accumulation of deleterious mutations in asexual eukaryote genomes using molecular sequences. *Journal of Natural History*, **34**, 1719–1729.

Ohta T (1987) Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*, **26**, 1–6.

Pala I, Schartl M, Brito M, Malta Vacas J, Coelho MM (2010) Gene expression regulation and lineage evolution: the North and South tale of the hybrid polyploid *Squalius alburnoides* complex. *Proceedings of the Royal Society of London B: Biological Sciences*, **277**, 3519–3525.

Paland S, Lynch M (2006) Transitions to asexuality result in excess amino acid substitutions. *Science*, **311**, 990–992.

Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PLoS One*, **6**, e22594.

Renny-Byfield S, Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany*, **101**, 1711–1725.

Riser NW (1994) The morphology and generic relationships of some fissiparous Heteronemertines. *Proceedings of the Biological Society of Washington*, **107**, 548–556.

Rogers AD, Thorpe JP, Gibson R (1995) Genetic evidence for the occurrence of a cryptic species with the littoral nemerteans *Lineus ruber* and *L. viridis* (Nemertea: Anopla). *Marine Biology*, **122**, 305–316.

Romiguier J, Figuet E, Galtier N *et al.* (2012) Fast and robust characterization of time-heterogeneous sequence evolutionary processes using substitution mapping. *PLoS One*, **7**, e33852.

Romiguier J, Gayral P, Ballenghien M *et al.* (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, **515**, 261–263.

Schaefer I, Domes K, Heethoff M *et al.* (2006) No evidence for the 'Meselson effect' in parthenogenetic oribatid mites (Oribatida, Acari). *Journal of Evolutionary Biology*, **19**, 184–193.

Schurko AM, Neiman M, Logsdon JM Jr (2009) Signs of sex: what we know and how we know it. *Trends in Ecology & Evolution*, **24**, 208–217.

Schwander T, Henry L, Crespi BJ (2011) Molecular evidence for ancient asexuality in *Timema* stick insects. *Current Biology*, **21**, 1129–1134.

Sémon M, Wolfe KH (2007) Consequences of genome duplication. *Current Opinion in Genetics & Development*, **17**, 505–512.

Simon JC, Delmotte F, Rispe C, Crease T (2003) Phylogenetic relationships between parthenogens and their sexual relatives: the possible routes to parthenogenesis in animals. *Biological Journal of the Linnean Society*, **79**, 151–163.

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research*, **19**, 1117–1123.

Small KS, Brudno M, Hill MM, Sidow A (2007) Extreme genomic variation in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5698–5703.

Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature*, **415**, 1022–1024.

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

Struck TH, Fisse F (2008) Phylogenetic position of Nemertea derived from phylogenomic data. *Molecular Biology and Evolution*, **25**, 728–736.

Sundberg P, Saur M (1998) Molecular phylogeny of some European heteronemertean (Nemertea) species and the monophyletic status of *Riseriellus*, *Lineus*, and *Micrura*. *Molecular Phylogenetics and Evolution*, **10**, 271–280.

Tsagkogeorga G, Cahais V, Galtier N (2012) The population genomics of a fast evolver: High levels of diversity, functional constraint, and molecular adaptation in the tunicate *Ciona intestinalis*. *Genome Biology and Evolution*, **4**, 740–749.

Tucker AE, Ackerman MS, Eads BD, Xu S, Lynch M (2013) Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15740–15745.

Wright S (1938) The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences of the United States of America*, **24**, 253–259.

Yi S, Ellsworth DL, Li WH (2002) Slow molecular clocks in Old World monkeys, apes, and humans. *Molecular Biology and Evolution*, **19**, 2191–2198.

Zimmer C (2009) On the origin of sexual reproduction. *Science*, **324**, 1254–1256.

Appendix I: A population genetic model of hybridization followed by asexual reproduction and the origin of truly private polymorphisms in *L. pseudolacteus*

Let us assume that at time 0, a triploid hybrid (*L. pseudolacteus*) is formed by fusion of two haplomes from diploid parent p_1 (*L. sanguineus*) with one haplome from diploid parent p_2 (*L. lacteus*_A). The p_1 and p_2 parents belong to different species, of effective population sizes N_1 and N_2, respectively, which are assumed to be constant in time. The parental and hybrid species then evolve during t generations. The hybrid species reproduces asexually, in absence of gene conversion between alleles and gene copy loss. Both the mutation rate, μ, and the generation time are assumed to be constant in time and among species. Only neutral mutations are considered, and multiple mutations at the same site are neglected.

Three distinct process are expected to contribute to the occurrence of truly private polymorphisms in present-day hybrid individuals, that is alleles segregating in the hybrid lineage but not observed in either of the two parental species: (i) new mutations having occurred in the hybrid lineage, (ii) ancestral polymorphism carried by p_1 and subsequently lost in the p_1 species and (iii) ancestral polymorphism carried by p_2 and subsequently lost in the p_2 species. Let H_{new}, H_{anc1} and H_{anc2} be the respective contribution of these three processes to truly private heterozygosity in present-day hybrids.

New mutations in the hybrid lineage

Under asexual reproduction in absence of gene conversion, gene copy loss and multiple mutations, lineages evolve independently from each other, and heterozygosity is never lost. The amount of newly created heterozygosity in t generations in a particular triploid hybrid lineage is therefore simply equal to:

$$H_{new} = 3\mu t \qquad \text{(eqn 1)}$$

Mutations inherited from p_1 and subsequently lost in the p_1 species

Assuming that the ancestral population was at mutation–drift equilibrium, the expected amount of heterozygosity initially inherited by the hybrid from the p_1 parent is $H_1(0)=4N_1\mu$. The per-generation rate of loss of heterozygosity by genetic drift in the p_1 species is $1/2N_1$, so that we have:

$$H_1(t) = H_1(0)(1 - 1/2N_1)^t \approx H_1(0)(1 - t/2N_1) \quad \text{(eqn 2)}$$

Rearranging, we obtain:

$$H_{anc1} = H_1(0) - H_1(t) = H_1(0)(t/2N_1) = 2\mu t \quad \text{(eqn 3)}$$

The expected amount of ancestral heterozygosity lost in the p_1 species, and therefore resulting in truly private polymorphism in the hybrid, is found to be independent from N_1. This occurs because the rate of heterozygosity loss is inversely related to the ancestral level of heterozygosity if N_1 is constant.

Mutations inherited from p_2 and subsequently lost in the p_2 species

The same rationale applies to parental species p_2, but as the hybrid inherits only a half of the ancestral diversity of p_2, the contribution to present-day truly private polymorphism in hybrids is half that of parent p_1:

$$H_{anc2} = H_{anc1}/2 = \mu t \quad \text{(eqn 4)}$$

From eqns (1), (3) and (4), we deduce that the expected proportion of present-day truly private polymorphisms in hybrid individuals having originated from new mutations in the hybrid lineage, $p_{new} = H_{new}/(H_{new} + H_{anc1} + H_{anc2})$, is equal to 0.5. This prediction is independent of the effective population size of the parental and hybrid species, but relies on the assumption of constant mutation rate and effective population size in time, and absence of gene conversion and gene copy loss in the hybrid lineage. If gene conversion and/or gene copy loss was at work in the hybrid lineage, then p_{new} would reach values higher than ½, as ancestral polymorphisms, which are older, are more likely than new mutations to have experienced a gene conversion/gene copy loss event at time t.

Missed shared polymorphisms

The above piece of theory is about heterozygosity gain/loss, whereas data analysis was mainly conducted based on SNP numbers. To connect theory with data, we need to account for the probability of SNP detection, which depends on sample size. In particular, a proportion of 'truly private' SNPs might actually correspond to shared polymorphisms that were missed due to limited sampling of the parental species – especially *L. lacteus*_A ($n = 5$).

Consider a mutant allele at frequency f in parental population 2 (*L. lacteus*_A) and absent from parental population 1 (*L. sanguineus*). With probability f, this allele will contribute a SNP in the hybrid (*L. pseudolacteus*), and assuming panmixia, the probability that this allele is undetected in the parental population based on a sample of n diploid individuals is $(1-f)^{2n}$. The expected ratio of undetected to detected shared SNPs is therefore:

$$q = \frac{\int h(f) f (1-f)^{2n} df}{\int h(f) f \left(1 - (1-f)^{2n}\right) df}$$

where $h(f)$ is the probability density of f. Assuming that $h(f)$ is proportional to $1/f$, as expected in a Wright–Fisher population (Wright 1938), the integrals can be calculated analytically and the ratio q is found to equal $1/2n$, that is $q = 0.1$ when $n = 5$. This implies that the number of *L. pseudolacteus* SNPs falsely assigned to the 'truly private' category while actually shared with *L. lacteus*_A equals approximately 10% of the SNPs assigned to the 'shared with *L. lacteus*_A' category – that is 80 SNPs. A similar calculation in *L. sanguineus* indicates that an additional ~67 SNPs assigned to the 'truly private' category are actually shared with *L. sanguineus*. The total number of missed shared polymorphisms is therefore estimated to represent ~40% of the 'truly private' category.

N.B. and N.G. designed the research project. N.B., N.G., E.E.Z., J.L.N., F.A.F.A. and J.B. provided biological material. J.L.N., F.A.F.A. and J.B. provided taxonomic expertise. E.F., E.E.Z. and M.B. generated the data. S.L.A.V., E.F., N.B. and N.G. analysed the data. S.L.A.V., N.B. and N.G. wrote the manuscript.

Data accessibility

Data sets are available from the NCBI/SRA database. Accession numbers are provided in Table S1 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Mitochondrial DNA phylogeny of *Lineus* nemerteans.

Fig. S2 Across-contigs distribution of observed heterozygosity in four *Lineus* species.

Fig. S3 Site frequency spectra in sexual (top) and fissiparous (bottom) *Lineus* species.

Fig. S4 Principal component analysis on the matrix of individual genotypes at 2531 SNPs.

Table S1 The sampled species and individuals.