

RESEARCH ARTICLE

Open Access



# Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): conserved genome structure, and ongoing low frequency recombination

Shanshan Dong<sup>1,2</sup>, Chaoxian Zhao<sup>1,3</sup>, Shouzhou Zhang<sup>1</sup>, Li Zhang<sup>1</sup>, Hong Wu<sup>2</sup>, Huan Liu<sup>4</sup>, Ruiliang Zhu<sup>3</sup>, Yu Jia<sup>5</sup>, Bernard Goffinet<sup>6</sup> and Yang Liu<sup>1,4\*</sup> 

## Abstract

**Background:** In contrast to the highly labile mitochondrial (mt) genomes of vascular plants, the architecture and composition of mt genomes within the main lineages of bryophytes appear stable and invariant. The available mt genomes of 18 liverwort accessions representing nine genera and five orders are syntenous except for *Gymnomitrium concinnatum* whose genome is characterized by two rearrangements. Here, we expanded the number of assembled liverwort mt genomes to 47, broadening the sampling to 31 genera and 10 orders spanning much of the phylogenetic breadth of liverworts to further test whether the evolution of the liverwort mitogenome is overall static.

**Results:** Liverwort mt genomes range in size from 147 Kb in Jungermanniales (clade B) to 185 Kb in Marchantiopsida, mainly due to the size variation of intergenic spacers and number of introns. All newly assembled liverwort mt genomes hold a conserved set of genes, but vary considerably in their intron content. The loss of introns in liverwort mt genomes might be explained by localized retroprocessing events. Liverwort mt genomes are strictly syntenous in genome structure with no structural variant detected in our newly assembled mt genomes. However, by screening the paired-end reads, we do find rare cases of recombination, which means multiple concurrent genome structures may exist in the vegetative tissues of liverworts. Our phylogenetic analyses of the nuclear encoded double strand break repair protein families revealed liverwort-specific subfamilies expansions.

**Conclusions:** The low repeat recombination level, selection, along with the intensified nuclear surveillance, might together shape the structural evolution of liverwort mt genomes.

**Keywords:** Bryophytes, recombination, structural evolution, repeats, introns

## Background

Mitochondrial (mt) genomes of vascular plants are highly variable in size (i.e., from ca. 66 Kb in *Viscum scurruloideum* [1] to 11.3 Mb in *Silene conica* [2]), and in gene content (i.e., 13 to 64 [3], excluding duplicated genes and open reading frames (ORFs)). By contrast, the mitogenomes within the three bryophyte lineages are conserved in terms of gene content, and exhibit rather

narrow size variation [4]. The 40 mt genomes from 29 moss genera (e.g., [5, 6]) are typically smaller (101–141 Kb, median ~ 107 Kb) than those of other land plants, and constantly encode 40 protein-coding genes (PCGs), 24 tRNAs, and 3 rRNAs. The mt genomes of hornworts sampled from four genera [7–9] are larger in size (185–242 Kb) but smaller in gene content (21–23 PCGs, 18–23 tRNAs, 3 rRNAs). The 18 mt genomes from seven liverwort genera [10–14] are intermediate in size between those of mosses and hornworts (142–187 Kb, median ~ 164 Kb) and may comprise more genes (i.e., 39–42 PCGs, 25–27 tRNAs, and 3 rRNAs).

\* Correspondence: [yang.liu0508@gmail.com](mailto:yang.liu0508@gmail.com)

<sup>1</sup>Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China

<sup>4</sup>BGI-Shenzhen, Shenzhen 518083, China

Full list of author information is available at the end of the article



Vascular plant mt genomes contain a variable set of introns, ranging in numbers from three (*Viscum album* [3]) to 37 (*Selaginella moellendorffii* [15]), with an average of 21. Bryophyte mt genomes tend to hold more introns than those of vascular plants, i.e., 28–38 (average 33) in hornworts [7], 26 or 27 in mosses [5], and 23–30 (average 28) in liverworts [13]. Each major land plant lineage contains a number of unique mt introns, and not a single intron is shared across all mt genomes of land plants [16]. As a result, the intron content is conserved within but not among land plant lineages [13, 16], suggesting that vast intron gains and losses happened during the early evolution of land plants, and conservative evolution maintained their stable content in the descendant lineages. Although each bryophyte group appears to hold a stable set of introns that parallels their conserved evolution of mt genomes of overall structure, recent small-scale mitogenomic studies provided evidences for distinct intron losses in leafy liverworts, and suggested retroprocessing as the likely causes [13]. Comprehensive studies with expanded taxon samplings are still needed to provide the framework for reconstructing the evolution of introns in liverwort mt genomes, and to assess the underlying causes for their variation in number.

Mt genomes of vascular plants hold abundant repeated sequences, including some large repeats (> 1000 bp), more medium sized repeats (100–1000 bp), and numerous small repeats (50–100 bp). Increased repeat length and identity facilitate intragenomic recombinations [16], and may hence account for the fluid genome structure of vascular mt genomes. In contrast, the mt genomes of the three bryophyte lineages usually have fewer repeated sequences and are structurally more stable. The mt genome of mosses contains only few and moreover small repeated sequences and is hence structurally nearly static [5]. Hornworts contain relatively more repeated sequences, which may explain the few rearrangements distinguishing the four assembled mitogenomes [7]. Finally, the mt genome of liverworts contains repeated sequences of intermediate abundance and size between those of mosses and hornworts [10]. The structure of their mt genome is highly conserved except for that of *Gymnomitrium concinnatum* [11], for which two inversions are needed to restore collinearity with the other liverwort mt genomes.

The nuclear encoded double-strand break repair (DSBR) proteins are involved in the suppression of the recombination between repeated DNA sequences [17]. Mutation of DSBR proteins in model organisms, such as *Arabidopsis* and *Physcomitrella*, often results in increased rearrangements between repeated sequences in organellar genomes that can impact the function of plastids and/or mitochondria [18–21]. The evolution of plant mt genome structure is therefore shaped by intrinsic parameters, such as repeat-mediated recombination, and extrinsic ones, such as

nuclear DSBR proteins [17]. The structure of the mt genome of bryophytes is considered highly conserved and stable [5] but the possible underlying causes such as the repeat recombination level and the evolution of the nuclear encoded DSBR genes remain unexplored. At present, only representatives of five of the currently recognized 15 orders of liverworts have seen their mitogenome assembled [22]. Here we broadened the sampling to exemplars of 10 orders to 1) test whether the mt gene content and 2) the structure of the mt genome is conserved across a broader phylogenetic breadth of liverworts; 3) investigate the repeat recombination rates in liverwort mitochondrion; 4) explore the possible intrinsic and extrinsic factors responsible for the structural evolutionary pattern of liverwort mt genomes.

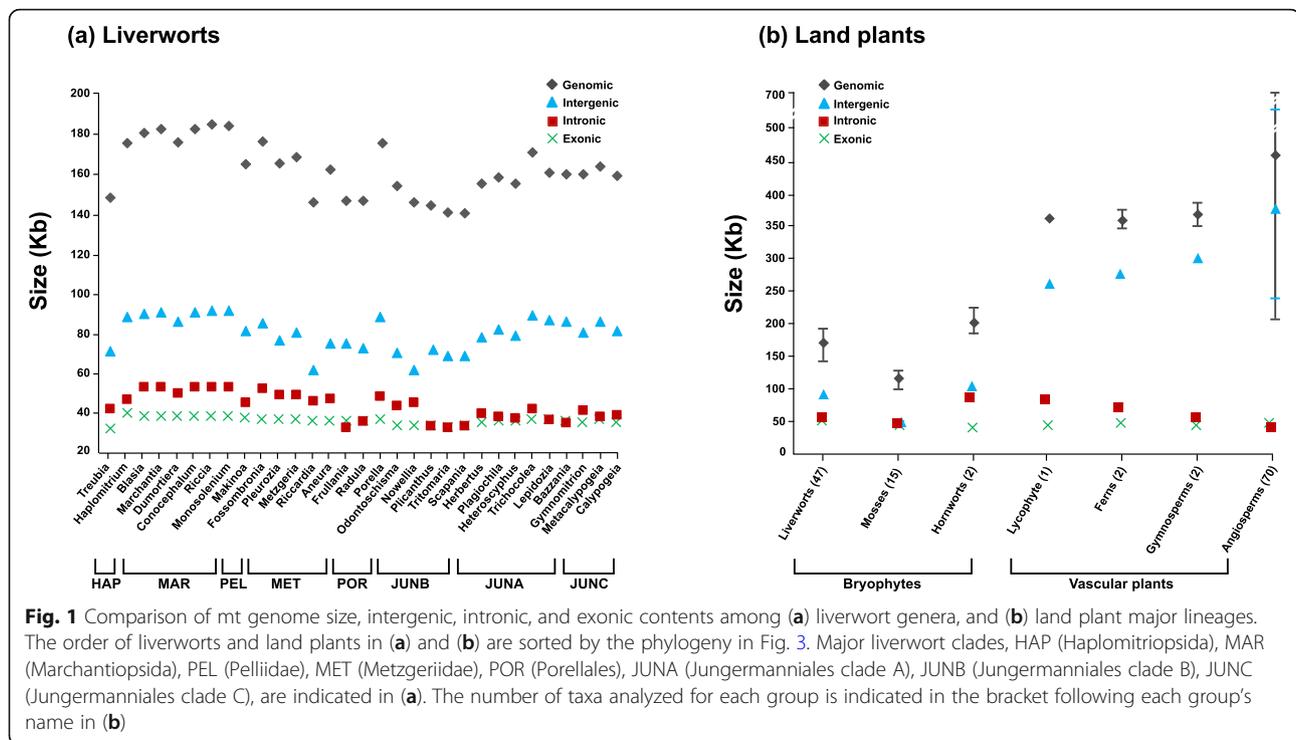
## Results and Discussion

### Liverwort mitogenomes are small in size but abundant in genes

Assemblies of high throughput sequences from total DNA extracts yielded complete circular mt genomes for 29 liverworts, representing 27 genera, 23 families, and 10 orders, of which five orders were newly sampled for their mt genomes. The sequencing depth ranges from 98 to 756 × (Additional file 1: Table S1). Including the published mt genomes from 13 species, complete liverwort mt genomes are thus available from 37 species, 31 genera, and 10 orders (Additional file 2: Table S2).

On average, liverwort mt genomes are smaller (~ 167 Kb) than those of vascular plants (average, ~ 476 Kb) or hornworts (average, ~ 197 Kb), but larger than those of mosses (average, ~ 107 Kb). Among liverworts, complex thalloids hold the largest (average, ~ 185 Kb) mt genomes, about 1.3 times the size of those of the Jungermanniopsida clade B that have the smallest mt genomes (average, ~ 147 Kb). Thus, variation in liverwort mitogenome size is much narrower compared to that in angiosperms (i.e., ~ 200-fold range). The trend in the evolution of mitogenome size during the diversification of liverworts is ambiguous (Fig. 1a), compared to a distinct increase during the evolution of land plants (Fig. 1b). The rather stable exome is consistently the smallest component (~ 23%) of liverwort mt genomes (~ 37 Kb). Consequently, changes in total mitogenome size in liverworts are shaped by the variation in intron content and, as shown in other land plants, intergenic spacer size (Fig. 1b).

The mitochondrial genome of liverworts spanning a broad phylogenetic breadth (Additional file 3: Figure S1) holds a nearly constant and large set of PCGs (39–43, Additional file 4: Figure S2). *Treubia* possesses the smallest set of mt PCGs (i.e., 39) among liverworts, and lacks all Cytochrome c genes (*ccmB*, *ccmC*, *ccmFN* and *ccmFC*), a characteristic first reported by Liu et al. [23] and confirmed here based on a new accession. A complete set of Cytochrome c genes is also lacking in the mitogenome of



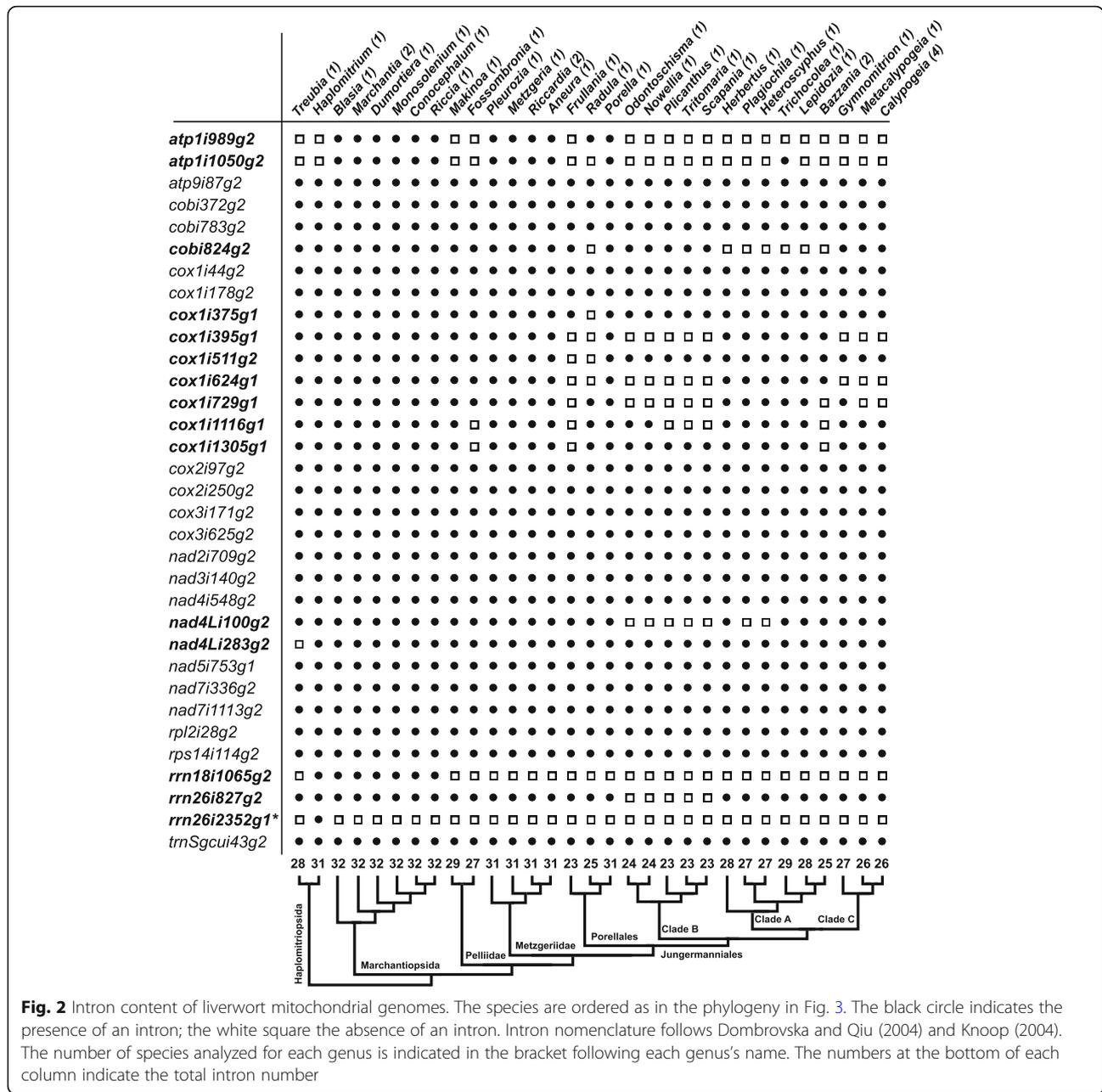
hornworts [7], lycophytes [24], and algae [16]. The *nad7* gene is widespread among algae, mosses, and vascular plants, but is pseudogenized in liverworts, except in the Haplomitriopsida (*Haplomitrium* and *Treubia*) that hold the intact and potentially functional *nad7*, lending support to the hypothesis of the earliest divergence of this lineage in the diversification of liverworts [25]. Ribosomal protein genes have been transferred from the mitochondria to the nucleus many times during the evolution of angiosperms [26]. By contrast, all liverwort mt genomes share an identical set of ribosomal protein genes, including *rps8* and *rps10* that rarely occur in other land plant mt genomes (Additional file 5: Table S3). Liverwort mt genomes also encode a rich set of RNA genes among land plants (Additional file 6: Figure S3), including three rRNA genes (*rrn5*, *rrn18* and *rrn26*) and 25–27 tRNA genes (i.e.,  $\pm$  *trnRucg* and *trnTggu*). Among land plants (Additional file 5: Table S3), *trnRucg* is restricted to the complex thalloid liverworts, and *trnTggu* occurs in only few liverwort lineages (i.e., *Blasia*, *Marchantia* and *Haplomitrium*), as well as in mosses and hornworts. Overall, the gene content of mitogenomes seems to be highly conserved during the evolution of liverworts, which spans for at least 400 million years [27].

#### Intron content of liverwort mitogenomes is much variable than that of mosses

Liverwort mt genomes hold a rich intron set, ranging from 23 introns in some leafy species to 32 in all the complex thalloid species, with an average of 28 introns (Fig. 2).

Species of Haplomitriopsida and Jungermanniopsida possess an average of 29 and 25 introns, respectively. Simple thalloid Pelliidae contains an average of 28 introns, whereas Metzgeriidae hold an average of 31 introns. Leafy Jungermanniidae show most cases of intron losses, especially in the mt genomes of *Frullania*, *Radula*, clade B and C of Jungermanniales (Fig. 2). Among the 17 genes disrupted by at least one intron, six (i.e., *atp1*, *cob*, *cox1*, *nad4L*, *rrn18*, and *rrn26*) vary in their intron content (Fig. 2). The three introns of *cob* gene are commonly present in liverworts, except for 3' end intron *cobi824g2* that is absent from *Radula* and clade A of Jungermanniales. The nine introns of *cox1* gene are shared by all thalloid liverworts, and the first three at the 5' end (*cox1i44g2*, *cox1i178g2*, and *cox1i375g1*) by almost all liverworts, but the remaining six variably lost among leafy liverworts. The respiratory chain complex I (or *nad*) genes possess a stable intron content, except for *nad4L*, for which one intron (*nad4Li100g2*) is lacking in a few Jungermanniales (clade B, and two species from clade A: *Plagiochila*, *Heteroscyphus*), and another one (*nad4Li283g2*) in *Treubia*. The liverwort unique *rrn18* group II intron (*rrn18i1065g2*) is present only in *Haplomitrium* and complex thalloids. Two introns were found in the ribosomal gene *rrn26*, and intron *rrn26i827g2* is present in all liverworts but clade B of the Jungermanniales, and *rrn26i2352g1*, a group I intron, is here newly reported and only observed in *Haplomitrium*.

Based on our phylogeny reconstructed from a concatenated nucleotide sequences of 41 mitochondrial



protein-coding genes (Fig. 2), the early diverging lineages Haplomitriopsida and Marchantiopsida tend to hold a more complete set of introns compared to the derived clades (especially clade B of Jungermanniales) that seem to have undergone parallel reduction in intron content. The alignments (Additional file 7: Figure S4; Additional file 8: Figure S5) of the four protein coding genes with intron number variations and RNA editing site distributions in liverworts support localized retroprocessing events [13] as the most possible causes for the intron loss phenomenon observed in liverworts. First, all exons are intact and the lost introns appear to be precisely cut off from the splicing site in liverwort

mt PCGs. Moreover, liverwort intron losses are more frequent toward the 3' end of genes (i.e., *cob*, *cox1*), which is a strong characteristic of retroprocessing induced intron losses [28]. Furthermore, in some cases, the intron losses in liverworts seem to be accompanied by RNA editing site losses near the exon-intron splicing boundaries.

**Intergenic spacer variations mirror the genome size variations in liverwort mitogenomes**

Intergenic spacers compose the bulk of the mt genome of land plants, accounting, for example, for about 80% in vascular plant mt genome size [16]. Intergenic spacers

comprise repeated sequences [29], sequences transferred from plastid [30] or nuclear genomes [31], and DNA fragments horizontally transferred from foreign donors [32, 33]. In liverworts, the intergenic spacers constitute the largest portion of mt genomes (average, ~ 81 Kb, ~ 49%), matching the combined exonic (~ 37 Kb, ~ 24%) and intronic (~ 44 Kb, ~ 27%) components (Table 1). The pattern in spacer length variation parallels that of genome size variation in liverworts (Fig. 1a). On average the complex thalloid mt genomes hold the largest spacer component (~ 91 Kb or ~ 49%), followed by that of *Haplomitrium* (~ 89 Kb or ~ 50%). The average spacer size decreases to ~ 83 Kb (~ 48%) in Porellales, 79 Kb (~ 50%) in Porellales, and ~ 74 Kb (~ 45%) in Metzgeriidae. In the Jungermanniales, the average spacer size is relatively small but varies between 84 Kb (~ 51%), 69 Kb (~ 46%), and 82 Kb (~ 51%) in clade A, B, C, respectively (Additional file 2: Table S2).

The spacer region of liverwort mt genomes is composed of ORFs, pseudogene fragments, nuclear homologous sequences, dispersed repeated sequences, SSRs, and other non-coding sequences of unknown origin. Plastid-derived or horizontally transferred DNA sequences are seemingly lacking in liverwort mt genomes (Table 1), as they were in those of mosses [5]. Nuclear derived sequences generally make up the largest component with an average of 20%, ranging from 13% in *Haplomitrium* to over 30% in the Marchantiopsida. Considering that we used the nuclear genome of *Marchantia polymorpha*, the only available nuclear genome for liverworts, as the reference for the blast search, it is not surprising that *Marchantia polymorpha* mt spacers returned the highest percentage hit (31%) on homologous sequences. The fast evolving/long branch lineage *Haplomitrium* (13%) and some simple thalloid taxa (such as *Aneura* and *Riccardia*) show the lowest percentage (i.e., 15%) of homologous sequence hits. Generally, the total size of the nuclear homologous sequence content in liverworts (average, ~ 15.7 Kb) is comparable to that in mosses (average, ~ 14.7 Kb), but the percentage relative to the total spacer size is only half that of mosses (average, ~ 42% [5]), which reflects the smaller size of moss mt spacers. ORFs usually make up the second largest part of the liverwort mt spacers, with an average ratio of 16%, ranging from 12% in some Jungermanniales species to 26% in *Haplomitrium*. The ratio of ORF to spacer size is smallest in the Marchantiopsida (average, ~ 13%), followed by the Jungermanniopsida (average, ~ 15%), then simple thalloids (average, ~ 17%). The ORF content in liverworts is distinctively larger than in mosses (i.e., 5%). As in mosses, SSR sequences compose generally less than 1% of the combined spacer region in liverwort mt genomes. Whereas moss mitogenomes lack repeated sequences in their intergenic spacers, liverworts hold on average 4% of repeated sequences in their intergenic spacer regions, with

only ~ 2% in the Haplomitriopsida and ~ 7% in the Marchantiopsida. The relatively higher repeated sequence content might allow for higher potential for mt genome recombination, as a positive correlation between the number of repeated sequences and the gene rearrangements has been suggested [5].

#### Liverwort mitogenomes are conserved in gene order despite repeat mediated recombinations

Repeated sequences play an important role in plant mt genome structure stability since they can mediate intra-genomic homologous recombinations leading to inversions and translocations of genomic regions [1, 16, 34]. Except for the mitogenome of *Gymnomitrium concinnatum* of the Jungermanniales [11], all remaining available liverwort mt genomes (i.e., 46) share exactly the same gene order (Fig. 3). The mt genome of *G. concinnatum* needs two inversions to make it collinear with the other liverwort mt genomes, which is in stark contrast to the situation in vascular plants wherein any two mt genomes require on average 31 rearrangements to gain collinearity [5]. The stable mt genome structure of liverworts might suggest either very low recombination level in liverwort mitochondrion and/or intensified nuclear surveillance over repeat recombinations.

Compared to the compact moss mt genomes with only a few small repeats shorter than 100 bp, liverwort mt genomes are much inflated, containing on average 14 pairs of small repeated sequences of 50–100 bp, and 14 pairs of medium-sized repeats of 100–900 bp (Additional file 9: Table S4). We examined the recombination rates of all repeats (535 pairs, Additional file 10: Table S5) within the 50–250 bp range for all the 29 samples and found recombination evidence for 26 repeats from 16 species (Table 2), 40% of the liverwort species show no evidence of recombination. The mitogenomes of nine leafy liverworts appear to recombine more frequently than those of complex thalloids (three species) and simple thalloids (two species). The two representatives of the early diverging Haplomitriopsida have recombinants detected. The repeats actively involved in recombination have an average size of the 103 bp with ten repeats (38%) exceeding 100 bp in length. Recombination rates range from 0.67 to 60% with a median value of 4.45%. About two thirds of the recombination events were mediated by small repeats (i.e., shorter than 100 bp) with a median recombination rate of 11.46%. Repeat length and recombination rate are not positively correlated. About half (12 out of 26) of the repeat mediated recombinations cause gene order changes and direct repeat recombinations affect genes more often than their inverted counterparts (62.5% vs 30%). Most (eight out of 10) inverted and four out of 16 direct repeat recombinations cause gene order changes. These gene order changes could give rise to alternative

**Table 1** Statistics of 47 available liverwort mitochondrial genomes. Taxa are sorted alphabetically. All ratios are calculated to the percentage of spacer length unless otherwise indicated. ORF = open reading frame; SSR = simple sequence repeat

Taxon	GenBank No.	Total spacer size (bp) (% of whole genome)	ORF (%)	Repeats (%) <sup>a</sup>	SSR (%)	Plastid origin (%)	Nuclear origin (%) <sup>b</sup>	Pseudogene (%)
<i>Aneura pinguis</i>	NC_026901	75,732 (45.73)	20.87	2.70	0.0	0	15.00	6.4
<i>Aneura pinguis</i>	KR817582	75,139 (45.54)	19.60	2.91	0.1	0	15.10	5.7
<i>Aneura pinguis</i>	KU140427	74,612 (45.48)	21.70	2.33	0.1	0	15.00	6.2
<i>Aneura pinguis</i>	KY242386	74,865 (45.46)	20.38	2.45	0.1	0	16.10	5.9
<i>Aneura pinguis</i>	KY702722	76,316 (46.01)	20.30	2.38	0.2	0	14.60	6.1
<i>Aneura pinguis</i>	KY702723	77,158 (46.19)	19.62	2.73	0.0	0	15.30	6.3
<i>Aneura pinguis</i>	MK230925	77,098 (46.19)	20.79	2.73	0.0	0	15.37	6.3
<i>Bazzania japonica</i>	MK230926	86,267 (53.05)	16.42	3.82	0.4	0	17.72	6.2
<i>Bazzania tridens</i>	MK318534	86,304 (53.06)	16.07	3.93	0.4	0	17.70	6.2
<i>Blasia pusilla</i>	MK230927	90,173 (48.76)	10.94	6.38	0.0	0	28.30	7.3
<i>Calypogeia arguta</i>	NC_035978	79,802 (50.17)	8.56	1.94	0.1	0	17.10	5.7
<i>Calypogeia integristipula</i>	NC_035977	82,607 (50.66)	13.16	3.56	0.1	0	19.00	6.7
<i>Calypogeia neogaea</i>	NC_035980	82,896 (51.12)	14.71	3.47	0.1	0	19.30	6.5
<i>Calypogeia suecica</i>	NC_035979	82,723 (51.08)	15.28	3.53	0.2	0	17.70	6.6
<i>Conocephalum conicum</i>	MK230928	91,366 (49.08)	12.35	6.09	0.1	0	28.93	6.3
<i>Dumortiera hirsuta</i>	MK230929	86,433 (48.56)	15.28	5.36	0.2	0	28.21	6.9
<i>Fossombronia cristula</i>	MK230936	85,625 (47.66)	14.22	4.25	0.1	0	20.98	6.6
<i>Frullania orientalis</i>	MK230937	75,479 (50.53)	12.11	4.34	0.3	0	19.59	4.4
<i>Gymnomitrium concinnum</i>	NC_040132	80,799 (49.70)	11.92	3.93	0.1	0	18.20	5.3
<i>Haplomitrium mnioides</i>	MK230931	89,333 (49.96)	26.24	2.71	0.0	0	12.81	5.3
<i>Haplomitrium mnioides</i>	MK230932	89,333 (49.96)	26.24	2.71	0.0	0	12.81	5.3
<i>Herbertus ramosus</i>	MK230946	78,882 (49.82)	12.79	4.19	0.2	0	22.08	8.5
<i>Heteroscyphus zollingeri</i>	MK230947	79,048 (50.09)	12.92	2.52	0.3	0	16.37	6.4
<i>Lepidozia trichodes</i>	MK230954	87,006 (53.25)	20.83	3.53	0.2	0	15.78	6.7
<i>Makinoa crispata</i>	MK230958	81,459 (48.73)	19.35	6.61	0.1	0	24.83	7.3
<i>Marchantia paleacea</i>	NC_001660	91,942 (49.27)	14.04	7.61	0.1	0	30.90	7.0
<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i>	NC_037508	91,328 (49.05)	12.53	7.61	0.1	0	30.79	7.1
<i>Metacalypogeia alternifolia</i>	MK230953	86,713 (51.99)	18.18	3.36	0.2	0	17.70	7.3
<i>Metzgeria leptoneura</i>	MK230956	80,637 (47.11)	18.08	3.70	0.1	0	18.26	6.6
<i>Monosolenium tenerum</i>	MK230931	92,474 (49.29)	15.15	6.82	0.1	0	29.42	6.8
<i>Monosolenium tenerum</i>	MK230932	92,476 (49.29)	15.15	6.80	0.1	0	29.42	6.8
<i>Nowellia curvifolia</i>	MK230952	61,971 (41.83)	11.66	2.93	0.1	0	14.34	6.7
<i>Odontoschisma grosseverrucosum</i>	MK230951	70,494 (45.28)	17.43	3.08	0.2	0	16.56	6.6
<i>Plagiochila subtropica</i>	MK230948	82,736 (51.26)	14.44	3.21	0.3	0	17.17	6.5
<i>Pleurozia purpurea</i>	MK230956	77,352 (45.90)	16.28	4.20	0.1	0	20.48	7.3
<i>Pleurozia purpurea</i>	NC_013444	77,347 (45.90)	15.81	4.19	0.1	0	20.50	7.3
<i>Plicanthus hirtellus</i>	MK230959	72,309 (49.32)	16.99	4.38	0.2	0	19.87	6.8
<i>Porella plumosa</i>	MK230938	88,966 (49.80)	16.81	5.63	0.1	0	21.08	6.6
<i>Radula japonica</i>	MK230939	73,160 (48.90)	20.68	4.42	0.3	0	15.13	5.7
<i>Riccardia latifrons</i>	MK230955	61,901 (41.62)	16.47	2.30	0.2	0	15.86	6.0
<i>Riccardia planiflora</i>	MK318535	62,537 (41.87)	18.13	1.89	0.11	0	14.90	5.1
<i>Riccia cavernosa</i>	MK230934	92,483 (49.23)	12.37	6.77	0.1	0	29.45	6.4

**Table 1** Statistics of 47 available liverwort mitochondrial genomes. Taxa are sorted alphabetically. All ratios are calculated to the percentage of spacer length unless otherwise indicated. ORF = open reading frame; SSR = simple sequence repeat (Continued)

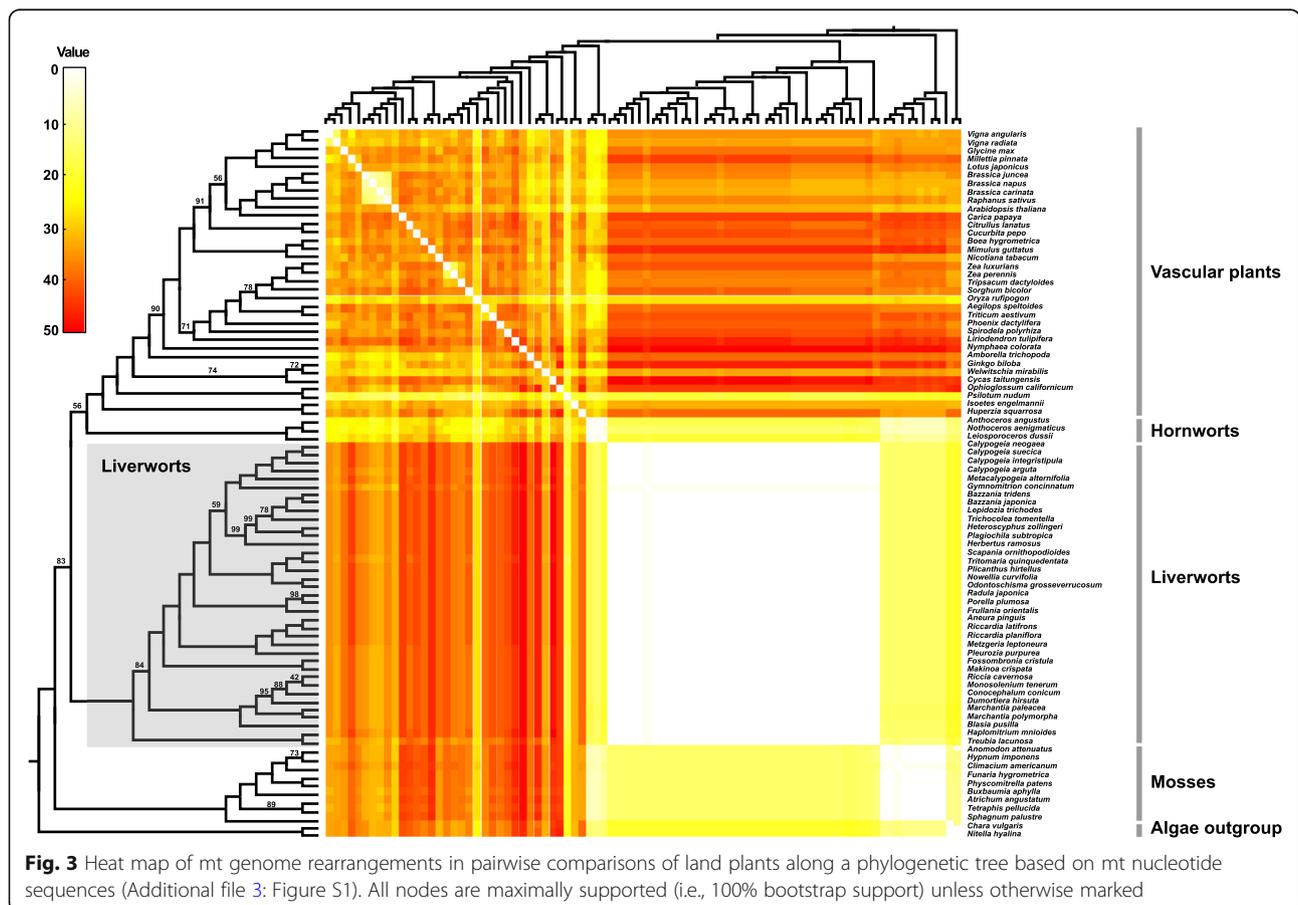
Taxon	GenBank No.	Total spacer size (bp) (% of whole genome)	ORF (%)	Repeats (%) <sup>a</sup>	SSR (%)	Plastid origin (%)	Nuclear origin (%) <sup>b</sup>	Pseudogene (%)
<i>Scapania ornithopodioides</i>	MK230950	68,944 (48.22)	11.81	3.40	0.1	0	17.42	7.3
<i>Treubia lacunosa</i>	NC_016122	71,724 (47.45)	16.60	2.32	0.0	0	18.55	6.1
<i>Treubia lacunosa</i>	MK230943	72,567 (47.75)	15.50	2.26	0.0	0	18.60	6.1
<i>Trichocolea tomentella</i>	MK230949	89,913 (51.74)	16.73	3.62	0.8	0	18.75	6.3
<i>Tritomaria quinquedentata</i>	MG640570	69,393 (48.69)	15.19	3.10	0.2	0	21.10	6.2

<sup>a</sup>Repeated sequences were detected by BLASTN, and sequences with similarity higher than 85% and aligned length longer than 50 bp were considered as repeat  
<sup>b</sup>Putative nuclear-derived sequences were detected by blasting the whole-intergenic spacer of each species against the *Marchantia polymorpha* nuclear genome; blast hits with E value 0.00001 were summarized. For length calculation, overlapping regions in mitochondrial spacers were removed

genome conformations, which, if they occurred in germ cells would be passed on to the offspring [35].

Although empirical studies suggest that repeats longer than 50 bp and with an identity above 85% may mediate recombinations [17, 34], the recombination activity of repeats is actually positively correlated with the length of repeated sequences, with small repeats (< 100 bp) rarely inducing recombinations [36, 37]. The detection of repeat recombinations for two thirds of liverwort species with an average of two active repeats per species might indicate repeat recombinations occurred, but in low

frequencies. Considering the average of five repeats (per species) longer than 250 bp not investigated for recombinations, it is very likely that some of these larger repeats (250–900 bp) may also allow for recombination. As recombination and structure fluidity are supposed to be positively correlated [1, 4, 5], the stable mt genome structure across liverwort diversity that spans over than 400 myr [27] is surprising given that liverwort mt genomes indeed recombine and alternative genome conformations coexisted. The apparent paradox of structural stasis of the mitogenome during the evolution



**Fig. 3** Heat map of mt genome rearrangements in pairwise comparisons of land plants along a phylogenetic tree based on mt nucleotide sequences (Additional file 3: Figure S1). All nodes are maximally supported (i.e., 100% bootstrap support) unless otherwise marked

**Table 2** Recombination frequency of repeats within 50–250 bp range for 16 liverwort mitogenomes

Species	Length (bp)	Identity (%)	Direction	Position	Reads support master circle conformation	Reads support alternative conformation	Gene order state after recombination	Genes affected
<i>Bazzania japonica</i>	51	94.12	direct	142,103–142,153	30 (68.18%)	14 (31.82%)	unchanged	<i>trnSUGA</i>
				141,693–141,743				
<i>Conocephalum conicum</i>	66	87.88	direct	70,639–70,704	39 (84.78%)	7 (15.22%)	unchanged	<i>atp8</i>
				70,053–70,118				
<i>Fossombronina cristula</i>	138	93.48	direct	85,167–85,304	148 (99.33%)	1 (0.67%)	unchanged	<i>cox2</i> and <i>cox3</i>
				84,355–84,492				
<i>Fossombronina cristula</i>	73	98.63	inverted	60,943–61,014	67 (57.76%)	49 (42.24%)	unchanged	–
				58,015–58,087				
<i>Fossombronina cristula</i>	95	86.32	direct	45,965–46,059	67 (60.91%)	43 (39.09%)	unchanged	–
				46,130–46,221				
<i>Fossombronina cristula</i>	93	87.1	inverted	61,247–61,339	86 (81.9%)	19 (18.1%)	unchanged	–
				58,038–58,130				
<i>Fossombronina cristula</i>	52	96.15	direct	55,570–55,619	80 (97.57%)	2 (2.43%)	unchanged	–
				55,443–55,494				
<i>Frullania orientalis</i>	73	98.63	direct	61,341–61,413	80 (98.73%)	1 (1.23%)	unchanged	–
				60,943–61,014				
<i>Frullania orientalis</i>	75	96	inverted	12,671–12,745	12 (92.31%)	1 (7.69%)	changed	–
				12,671–12,745				
<i>Haplomitrium mnioides</i> I	77	88.31	direct	87,784–87,860	48 (97.96%)	1 (2.04%)	changed	<i>cox3</i> and <i>trnSGCU</i>
				40,562–40,637				
<i>Haplomitrium mnioides</i> I	241	100	direct	140,823–141,063	95 (98.96%)	1 (1.04%)	changed	–
				132,233–132,473				
<i>Heteroscyphus zollingeri</i>	51	90.2	direct	14,195–14,245	35 (97.2%)	1 (2.8%)	unchanged	–
				14,085–14,135				
<i>Heteroscyphus zollingeri</i>	253	96.05	direct	66,972–67,223	27 (96.43%)	1 (3.57%)	changed	<i>rps7</i>
				21,339–21,591				
<i>Monosolenium tenerum</i> I	138	92.03	direct	86,651–86,787	94 (98.95%)	1 (1.05%)	unchanged	<i>cox2</i> and <i>cox3</i>
				85,840–85,977				
<i>Nowellia curvifolia</i>	120	100	inverted	88,658–88,777	31 (60.78%)	20 (39.22%)	changed	<i>cob</i>
				5529–5648				
<i>Nowellia curvifolia</i>	138	98.55	inverted	88,962–89,099	62 (78.48%)	17 (21.52%)	changed	<i>cob</i>
				5392–5529				

**Table 2** Recombination frequency of repeats within 50–250 bp range for 16 liverwort mitogenomes (Continued)

Species	Length (bp)	Identity (%)	Direction	Position	Reads support master circle conformation	Reads support alternative conformation	Gene order state after recombination	Genes affected
	64	93.75	inverted	14,0191–140,253	30 (96.77%)	1 (3.23%)	changed	–
<i>Odontoschisma grosseverrucosum</i>	52	94.23	inverted	14,0191–140,253 94,954–95,005 5637–5688	32 (40%)	48 (60%)	changed	<i>cob</i>
<i>Pleurozia purpurea</i>	140	98.57	direct	113,528–113,667 108,644–108,788	21 (95.46%)	1 (4.54%)	unchanged	<i>atp1</i>
<i>Plicanthus hirtellus</i>	54	98.15	inverted	112,052–112,105 84,436–84,489	31 (48.44%)	33 (51.56%)	changed	–
<i>Porella plumosa</i>	62	90.32	inverted	45,426–45,485 45,426–45,485	20 (95.24%)	1 (4.76%)	changed	–
<i>Radula japonica</i>	114	89.47	direct	68,120–68,227 67,321–67,434	49 (98%)	1 (2%)	unchanged	<i>cox2</i> and <i>cox3</i>
	189	99.47	direct	142,227–142,415 79,402–79,590	67 (98.53%)	1 (1.47%)	changed	–
<i>Riccia cavernosa</i>	77	87.01	inverted	110,110–110,185 110,110–110,185	66 (95.65%)	3 (4.35%)	changed	–
<i>Treubia lacunosa</i>	137	93.43	direct	72,748–72,884 72,001–72,137	82 (98.8%)	1 (1.2%)	unchanged	<i>cox2</i> and <i>cox3</i>
<i>Trichocolea tomentella</i>	63	92.06	direct	151,893–151,955 151,475–151,537	14 (82.35%)	3 (17.65%)	unchanged	<i>trnSUGA</i>

of liverworts despite evidence of ongoing recombination may be addressed in the following three contexts. First, as some authors reported the adaptive value of repeats and recombinations in plant mitochondrion [37], recombinations happen more frequently when the cells are in stress and/or under some environmental stimuli [38], it is likely that the low recombinations observed in liverworts might primarily happen in the old vegetative cells rather than in young differentiating cells and germ cells, therefore the liverwort progeny inherited the master circle conformation. Second, if repeat recombinations occurred in the reproductive cells, those recombinants with genes affected or genes missing might be selected against. Recent studies in *Drosophila* also suggested a generalizable mechanism for selection against deleterious alternative mt configurations: ATP selection after mitochondrion fragmentation can retain those mtDNA fragments that contain the 'correct' mt genome with the complete set of the genes [39]. Nevertheless, those alternative conformations with altered gene order but no genes affected may survive this ATP selection and might possibly lead to offspring with rearranged gene order. Finally, It is also possible that maintaining the organization of genes is essential to the transcription of polycistronic operons in liverworts and mosses [5], hence an identical gene order is selected across all liverwort lineages despite the existence of alternative genome conformations.

#### **Liverwort mt genomes might be under intensified nuclear surveillance**

The structural stability of liverwort mt genomes, in accordance with the remarkable structural conservatism of liverwort plastomes [40], might also be shaped by nuclear encoded DSBR proteins that suppress the error-prone ectopic recombinations across small direct repeats during DSBR in mitochondrion and plastid [41]. We here characterized six frequently reported DSBR gene families in the transcriptome assemblies of 125 land plant representatives (Additional file 11: Table S6). Functional studies have confirmed the mtDNA repair function of the moss orthologs in *Physcomitrella patens* for four of these gene families, *RecA*, *RecG*, *RecX*, and *MSH1*. In our phylogenetic analyses (Additional file 12: Figure S6), four DSBR gene families (*RecA*, *RecG*, *RecX*, and *OSB*) show notable liverwort-specific subfamily expansions, suggesting peculiar evolutionary pattern of DNA repair mechanisms in liverworts. The subcellular localization of these DSBR proteins (as predicted by TargetP) yielded ambiguous results for most of these DSBR proteins, including those expanded ones, suggesting either truncated protein sequences or incomplete query database. Although the subcellular localization (as predicted by TargetP), and the in vivo function of these expanded liverwort gene family members remain elusive,

we could not rule out the possibility that these liverwort specific expansions might also contribute to the mtDNA sequence stability of liverworts. Specifically, selective pressure might have driven the functional divergence of DSBR gene families under different evolutionary constraints, and that those diversified DSBR proteins may exhibit a wider range of characteristics and specificities, which possibly help intensify the DNA repair mechanisms in liverworts.

#### **Conclusions**

Based on the assembly of mitochondrial genomes for a widely expanded sample of liverworts spanning over 400 million years of evolution [27], we a) confirm that the mitogenome of liverworts is highly conserved in both gene content and gene order, while having a large and variable set of introns (23–32 introns) subject to a series of localized retroprocessing events, b) reveal ongoing intragenomic recombination mediated by small repeats and c) show liverwort specific DSBR subfamily expansions in four gene families. Although the mt genomes are relatively conserved in each of the three major bryophyte lineages (i.e., liverworts, mosses and hornworts), they differ in the content of repeat sequences and frequency of rearrangements, suggesting that the mechanisms maintaining mitogenome structure differ among these lineages. The conserved evolution of liverwort mt genomes might be explained by low recombination level, selection, and intensified nuclear surveillance.

#### **Methods**

##### **Materials and DNA extractions**

Wild samples of liverwort materials of 29 accessions were collected in field trips from China (Xizang, Sichuan, Yunnan, Fujian, and Guangzhou Provinces), Madagascar, the United States (Connecticut), Vietnam, and New Zealand (Additional file 1: Table S1). No specific permissions were needed on current study of these samples. These samples were studied and identified by Qiang He, and the voucher specimens were deposited in SZG (Herbarium of Shenzhen Fairy lake Botanical Garden, Shenzhen, China). Our sampling spreads across the liverwort phylogeny, representing 10 of the 15 orders of extant liverworts. Each sample was cleaned with distilled water and dried using lab paper. The clean shoots of individual samples were isolated under the dissecting microscope and used for DNA extractions using the modified CTAB methods [42]. The DNA quality and quantity of each sample were examined using 1% Agarose gel electrophoresis and Nanodrop 2000 spectrophotometer.

##### **Mitochondrial genome sequencing and assembly**

For genomic DNA sequencing, 1 µg high quality genomic DNA was sheared using the Covaris M220 (Woburn, MA,

USA), DNA fragments of 300–500 bp were used to generate sequencing libraries using the Illumina TruSeq™ DNA PCR-free library preparation kit (Illumina, CA, USA) following the manufacturer's instructions. The libraries were paired-end ( $2 \times 150$  bp) sequenced on an Illumina HiSeq 2000 sequencing platform at the WuXiNextCode (Shanghai, China). Approximately 10 Gb sequencing data were generated for each sample. The raw NGS data were trimmed and filtered for adaptors, low quality reads, undersized inserts, and duplicate reads using Trimmomatic [43]. The filtered reads for each species were de novo assembled using CLC Genomics Workbench v5.5 (CLC Bio, Aarhus, Denmark). All assembled contigs were blasted to the *Marchantia polymorpha* mt genome (GenBank accession: NC\_001660) to identify mt contigs. As the genomes of the three cellular compartments are significantly different in copy numbers and hence read coverage in sequencing [44], the read depth for mt contigs is distinctly lower than that of plastid contigs, but significantly higher than that of nuclear contigs [45]. For each species, every resulting mt contigs (usually with one to five per species) was first checked for read depth to ensure they are in a similar range, and then each of these mt contig was elongated at both ends in Geneious v10.0.2 (Biomatters, New Zealand) using Bowtie2 [46] till their ends overlapped with one another by at least 5000 bp. Altogether, 29 circular mt genomes were assembled for liverworts. Finally, the corresponding genomic reads were mapped back to the complete mitochondrial genomes to double check for sequencing depth (Additional file 1: Table S1) of the assembled mt genomes. All the 29 liverwort mt genomes newly assembled in this study received constant depth along the mt genomes.

#### Genome annotation and comparative analysis

The mt genomes were annotated following the steps described by Li et al. [9] and Xue et al. [8]. In summary, the protein-coding and rRNA genes were annotated by Blastn searches of the non-redundant database of National Center for Biotechnology Information (NCBI). The exact gene and exon/intron boundaries were further verified in Geneious v10.0.2 (Biomatters, New Zealand) by aligning orthologous genes from the published annotated liverwort mitochondrial genomes. The tRNA genes were annotated using tRNAscan-SE v2.0 [47]. Mitochondrial RNA editing sites were obtained from our previous study on liverwort organellar RNA editing [48]. We summarized and compared the structural evolution of all the formally published (as by June 2019) liverwort mt genomes (Additional file 2: Table S2) from the gff3 annotation files. As accessions from the same species/genus show very similar characters in mt genomes in all aspects, directly averaging the values across all organisms would bias towards better-represented lineages. Therefore, to avoid such biases we calculated the average

values under three different strategies, if multiple accessions of a species exist: (1) we average them first and then calculate the average value among species (e.g., genome/exon/intron length, and gene/intron number); (2) similarly for among genera, we averaged the values of species, and (3) for among clades, we averaged the value of each genus.

#### Repeats and recombination analysis

Repeat sequences are annotated and analyzed (Additional file 9: Table S4) using the python tool developed by Wynn & Christensen [37]. Considering at least 50 bp matches at both ends of the insert fragment and an average size of insert fragments being 350 bp in our sequencing libraries, we can estimate the recombination rate of all non-overlapping repeats between 50 bp and 250 bp. Altogether, we analyzed 534 repeats for the 29 liverwort mt genomes (Additional file 10: Table S5) following Dong et al. [36]. Specifically, for each repeat pair, we built four or eight reference sequences, each with 200 bp up- and down-stream of the two template sequences (original sequences), and the two (for repeat pair with identity = 100) or six (for repeat pair with identity < 100) recombined sequences (alternative configurations) constructed from the putative recombination products. All the recombined templates were blasted against the corresponding mt genome sequences to identify those on the genome. Then, we searched the reference sequences against the corresponding paired-end reads database, and counted the number of matching read pairs with a blast identity above 98%, and a hit coverage of at least 50 bp in both flanking regions of each repeat sequence. After that, the best matched read pairs for all the recombinants were extracted and mapped to the corresponding mt genomes using Bowtie2 [46] and the resultant bam file was visualized in Geneious v10.0.2 (Biomatters, New Zealand) to confirm the existence of the alternative configurations in reads.

#### Gene alignment and phylogenetic reconstruction

For the phylogenetic reconstruction, sequences of each of the 41 mitochondrial genes were aligned using TranslatorX [49], which first translates the nucleotide (nt) sequence into amino acids (aa) using the standard, universal genetic code, and then uses MAFFT [50] to create an amino acid alignment. The alignment is further trimmed for ambiguous portions by GBLOCKS [51] with the least stringent settings. The cleaned aa alignment is then used as a guide to generate the nt sequence alignment. The resultant nt alignments of individual genes were then combined as one dataset in Geneious v10.0.2 (Biomatters, New Zealand). The concatenated nt dataset was analyzed by the maximum likelihood (ML) method as implemented in the RAxML software [52] with codon-partitioned GTRGAMMA model. Branch

support for each internode was evaluated with 300 bootstrap approximation (Additional file 3: Figure S1). The mitochondrial phylogeny of liverworts is mostly congruent with that of the plastid phylogeny [53].

### Identification of DSBR proteins and phylogenetic analysis

We used HMMER [54] to conduct profile hidden Markov model (HMM) searches using the Pfam *RecA* (PF00154), *SSB* (PF00436), *Whirly* (PF08536), *RecX* (PF02631), *MutS\_V* (PF00488), *DEAD* and *Helicase\_C* (PF00270, PF00271) domains as queries to search the annotated proteins from 125 plant species (Additional file 11: Table S6) for characterization of the gene family members of *RecA*, *OSB*, *Why*, *RecX*, *MSH*, and *RecG*, respectively. For gene loci with multiple isoforms predicted, the primary isoform was used if primary isoform annotation is available; otherwise the longest protein was used. We considered sequences with Pfam domain identified by HMMER with an E-value of  $1e-6$ , and the hit alignment length of at least 50% of the domain length to be candidate proteins. Those candidate proteins from the HMMER search were further manually confirmed using the SMART [55] and Pfam [56] databases. The proteins sequences were filtered for redundancy using CD-hit [57] with a similarity threshold of 0.95, aligned using MATTF [50], and trimmed for ambiguous portions by GBLOCKS with the least stringent settings [51]. The final protein alignment were analyzed by the maximum likelihood (ML) method as implemented in the IQ-TREE software [58] with 1000 ultrafast bootstrap replicates. The subcellular localizations of these DSBR proteins were predicted using online TargetP 1.1 server [59].

### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-019-6365-y>.

**Additional file 1: Table S1.** Overview of the 29 newly generated liverwort mitochondrial genomes.

**Additional file 2: Table S2.** Comparison of the genome content of the 47 liverwort mitochondrial genomes.

**Additional file 3: Figure S1.** ML phylogram of selected land plants with an emphasis on liverworts based on a concatenated nucleotide data set.

**Additional file 4: Figure S2.** Distribution of mitochondrial protein genes in 31 liverwort genera.

**Additional file 5: Table S3.** Comparison of gene content and gene order from 88 selected land plant mitochondrial genomes.

**Additional file 6: Figure S3.** Distribution of mitochondrial RNA genes in 31 liverwort genera.

**Additional file 7: Figure S4.** RNA editing site variations and intron losses in the *cox1* gene as an exemplar.

**Additional file 8: Figure S5.** RNA editing site distributions on three gene alignments in a phylogenetic context.

**Additional file 9: Table S4.** Distribution of repeat sizes among liverwort mitochondrial genomes.

**Additional file 10: Table S5.** Repeat recombinations for 534 repeats within 50 to 300 bp range in liverwort mitochondria.

**Additional file 11: Table S6.** Taxa list for DSBR protein identification.

**Additional file 12: Figure S6.** Phylogenetic trees of DSBR protein sequences from 125 Viridiplantae taxa inferred by Iqtree.

### Abbreviations

aa: Amino acid; DSBR: Double strand break repair; ML: Maximum likelihood; Mt genome: Mitochondrial genome; mtDNA: Mitochondrial DNA; nt: Nucleotide; ORFs: Open reading frames; PCGs: Protein-coding genes; rRNAs: Ribosomal RNAs; SZG: Herbarium of Shenzhen Fairy lake Botanical Garden, Shenzhen, China; tRNAs: Transfer RNAs

### Acknowledgments

We are highly grateful to Dr. Qin Zuo and Hui Dong at the Shenzhen Fairy Lake Botanical Garden for assistance in acquiring liverwort materials. We wish to thank Dr. Qiang He and Qinghua Wang from the Institute of Botany (CAS) for the help in fieldwork and taxonomic consultation. We also thank David Glenny for providing fresh *Treubia* materials. We gratefully acknowledge the lab assistances by Yang Peng and Na Li at the Shenzhen FairyLake Botanical Garden.

### Authors' contributions

YL conceived and designed the study. YL, SD, CZ, RZ, and BG collected the liverwort materials. SD performed the experiments. SD and CZ prepared the figures. SD and YL carried out the analyses and drafted the manuscript. YL, HL, SZ, LZ and BG revised the manuscript. All authors have read and approved the final manuscript.

### Funding

This project is fund by the National Science Foundation of China (31470314, 31600171). The funding bodies did not participate in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

### Availability of data and materials

Newly generated mitochondrial genomes of 29 liverworts have been deposited in GenBank under the accession numbers MK230925–MK230962. The raw genomic NGS read data have been deposited in the Short Read Archive (SRA) database of NCBI under the study number SRP170808. Other supporting results are included within the article and its additional files.

### Ethics approval and consent to participate

Fieldwork was conducted in accordance with local legislation.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, China. <sup>2</sup>State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou 510642, China. <sup>3</sup>Department of Biology, School of Life Sciences, East China Normal University, Shanghai 200241, China. <sup>4</sup>BGI-Shenzhen, Shenzhen 518083, China. <sup>5</sup>State Key Laboratory of Systematic and Evolutionary Botany, Chinese Academy of Sciences, Institute of Botany, Beijing 100093, China. <sup>6</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3043, USA.

Received: 6 September 2019 Accepted: 2 December 2019

Published online: 09 December 2019

### References

1. Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A*. 2015;112(27):E3515–24.

2. Sloan DB, Muller K, McCauley DE, Taylor DR, Storchova H. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol.* 2012;196(4):1228–39.
3. Petersen G, Cuenca A, Møller IM, Seberg O. Massive gene loss in mistletoe (*Viscum, viscaceae*) mitochondria. *Sci Rep.* 2015;5(1):17588.
4. Liu Y, Wang B, Li L, Qiu YL, Xue J. Conservative and dynamic evolution of mitochondrial genomes in early land plants. *Genomics Chloroplasts Mitochondria Springer Neth.* 2012;35:159–74.
5. Liu Y, Medina R, Goffinet B. 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. *Mol Biol Evol.* 2014;31(10):2586–91.
6. Vigalondo B, Yang L, Draper J, Lara F, Garilletti R, Mazimpaka V, Goffinet B. Comparing three complete mitochondrial genomes of the moss genus *Orthotrichum* Hedw. *Mitochondrial DNA Part B Resour.* 2016;1(1):179–81.
7. Dong S, Xue JY, Zhang S, Zhang L, Wu H, Chen Z, Liu Y. Complete mitochondrial genome sequence of *Anthoceros angustus*: conservative evolution of the mitogenomes in hornworts. *Bryologist.* 2018;121(1):14–23.
8. Xue JY, Liu Y, Li L, Wang B, Qiu YL. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Curr Genet.* 2009;56(1):53–61.
9. Li L, Wang B, Liu Y, Qiu YL. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *J Mol Evol.* 2009;68(6):665–78.
10. Dong S, He Q, Zhang S, Wu H, Goffinet B, Liu Y. The mitochondrial genomes of *Bazzania tridens* and *Riccardia planiflora* further confirms conservative evolution of mitogenomes in liverworts. *Bryologist.* 2019;122:130–9.
11. Myszczyński K, Górski P, Ślipiko M, Sawicki J. Sequencing of organellar genomes of *Gymnomitrium concinnatum* (Jungermanniales) revealed the first exception in the structure and gene order of evolutionarily stable liverworts mitogenomes. *BMC Plant Biol.* 2018;18:321.
12. Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, et al. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. *J Mol Biol.* 1992;223(1):1–7.
13. Ślipiko M, Myszczyński K, Buczkowska-Chmielewska K, Bączkiewicz A, Szczecińska M, Sawicki J. Comparative analysis of four *Calypogeia* species revealed unexpected change in evolutionarily-stable liverwort mitogenomes. *Genes-Basel.* 2017;8(12):395.
14. Wang B, Xue J, Li L, Yang L, Qiu YL. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea*, reveals extremely conservative mitochondrial genome evolution in liverworts. *Curr Genet.* 2009;55(6):601–9.
15. Hecht J, Grewe F, Knoop V. Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. *Genome Biol Evol.* 2011;3:344–58.
16. Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: The genomics revolution. Heidelberg: Springer Vienna; 2012.
17. Marechal A, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* 2010;186(2):299–317.
18. Abdelnoor RV, Christensen AC, Mohammed S, Munoz-Castillo B, Moriyama H, Mackenzie SA. Mitochondrial genome dynamics in plants and animals: Convergent gene fusions of a *MutS* homologue. *J Mol Evol.* 2006;63:165–73.
19. Zaegel V, Guermann B, Le Ret M, Andres C, Meyer D, Erhardt M, Canaday J, Gualberto JM, Imbault P. The plant-specific ssDNA binding protein *OSB1* is involved in the stoichiometric transmission of mitochondrial DNA in *Arabidopsis*. *Plant Cell.* 2006;18:3548–63.
20. Shedje V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. Plant mitochondrial recombination surveillance requires unusual *RecA* and *MutS* homologs. *Plant Cell.* 2007;19:1251–64.
21. Marechal A, Parent JS, Veronneau-Lafortune F, Joyeux A, Lang BF, Brisson N. Whirly proteins maintain plastid genome stability in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2009;106:14693–8.
22. Söderström L, Hagborg A, Konrat MV, Bartholomew-Began S, Bell D, Briscoe L, Brown E, Cargill DC, Costa DP, Crandall-Stotler B, et al. World checklist of hornworts and liverworts. *Phytokeys.* 2016;59(59):1–828.
23. Liu Y, Xue JY, Wang B, Li L, Qiu YL. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: Dynamic and conservative evolution. *PLoS One.* 2011;6:e25836.
24. Grewe F, Herres S, Viehover P, Polsakiewicz M, Weisshaar B, Knoop V. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.* 2011;39(7):2890–902.
25. Groth-Malonek M, Knoop V. Bryophytes and other basal land plants: the mitochondrial perspective. *Taxon.* 2005;54(2):293–7.
26. Daley DO, Adams KL, Clifton R, Qualmann S, Millar AH, Palmer JD, Pratje E, Whelan J. Gene transfer from mitochondrion to nucleus: novel mechanisms for gene activation from *cox2*. *Plant J.* 2002;30(1):11–21.
27. Morris JL, Puttick MN, Clark JW, Edwards D, Kenrick P, Pressel S, Wellman CH, Yang Z, Schneider H, Donoghue PCJ. The timescale of early land plant evolution. *Proc Natl Acad Sci U S A.* 2018;115(10):E2274–83.
28. Cuenca A, Ross TG, Graham SW, Barrett CF, Davis JJ, Seberg O, Petersen G. Localized retroprocessing as a model of intron loss in the plant mitochondrial genome. *Genome Biol Evol.* 2016;8(7):2176–89.
29. Lilly JW, Havey MJ. Small, repetitive dnas contribute significantly to the expanded mitochondrial genome of cucumber. *Genetics.* 2001;159(1):317–28.
30. Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 2010;27(6):1436.
31. Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *Plant Cell.* 2011;23(7):2499–513.
32. Bock R. Witnessing genome evolution: experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annu Rev Genet.* 2017;51:1–22.
33. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchezpuerta MV, Munzinger J, Barry K, Boore JL, Zhang L, DePamphilis CW, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science.* 2013;342(6165):1468–73.
34. André C, Levy A, Walbot V. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* 1992;8:128–32.
35. Woloszyńska M. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. *J Exp Bot.* 2010;61(3):657–71.
36. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, Zhang L, Liu Y. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics.* 2018; 19(1):614.
37. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
38. Wallet C, Le Ret M, Bergdoll M, Bichara M, Dietrich A, Gualberto JM. The RECG1 DNA translocase is a key factor in recombination surveillance, repair, and segregation of the mitochondrial DNA in *Arabidopsis*. *Plant Cell.* 2015;27:2907–25.
39. Lieber T, Jeedigunta SP, Palozzi JM, Lehmann R, Hurd TR. Mitochondrial fragmentation drives selective removal of deleterious mtDNA in the germline. *Nature.* 2019;570:380.
40. Yu Y, Liu H, Yang J, Ma W, Pressel S, Wu Y, Schneider H. Exploring the plastid genome disparity of liverworts. *J Syst Evol.* 2019;57(4):382–94.
41. Alexandre M, Brisson N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* 2010;186(2):299–317.
42. Porebski S, Bailey LG, Bernard RB. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Report.* 1997;15(1):8–15.
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
44. Dierckxns N, Mardulyn P, Smits G. Novoplasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 2017;45:gw955.
45. Zhang T, Zhang X, Hu S, Yu J. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods.* 2011;7(1):38.
46. Langmead B. Fast gapped-read alignment with Bowtie2. *Nat Methods.* 2012; 9(4):357–9.
47. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016:gw413.
48. Dong S, Zhao C, Zhang S, Wu H, Mu W, Wei T, Li N, Wan T, Liu H, Cui J, et al. The amount of RNA editing sites in liverwort organellar genomes is correlated with the GC content and PPR protein diversity. *Genome Biol Evol.* 2019. <https://doi.org/10.1093/gbe/evz232>.
49. Abascal F, Zardoya R, Telford MJ. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* 2010;38(suppl\_2):W7–W13.
50. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005;33(2):511–8.

51. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56(4):564–77.
52. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006; 22(21):2688–90.
53. Yu Y, Yang J, Ma W, Pressel S, Liu H, Wu Y, Schneider H. Chloroplast phylogenomics of liverworts: a reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. *Cladistics.* 2019;0:1–10.
54. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39:29–37.
55. Letunic I, Peer B. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* 2017;46:D493–6.
56. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heeger A, Hetherington K, Holm L, Mistry J, et al. Pfam: the protein families database. *Nucleic Acids Res.* 2014;42:222–30.
57. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 2006;22(13):1658.
58. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
59. Emanuelsson O, Nielsen H, Brunak S, Heijne GV. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol.* 2000;300:1005–16.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

