





RESOURCE

The genome sequence of *Hirschfeldia incana*, a new Brassicaceae model to improve photosynthetic light-use efficiency

Francesco Garassino^{1,†} , Raúl Y. Wijfjes^{2,†,‡} , René Boesten^{1,†}, Francisca Reyes Marquez¹, Frank F. M. Becker¹, Vittoria Clapero^{1,†}, Iris van den Hatert², Rens Holmer², M. Eric Schranz³ , Jeremy Harbinson⁴, Dick de Ridder², Sandra Smit^{2,*} and Mark G. M. Aarts^{1,*} 

¹Laboratory of Genetics, Wageningen University & Research, Wageningen, Netherlands,

²Bioinformatics Group, Wageningen University & Research, Wageningen, Netherlands,

³Biosystematics Group, Wageningen University & Research, Wageningen, Netherlands, and

⁴Laboratory of Biophysics, Wageningen University & Research, Wageningen, Netherlands

Received 14 April 2022; revised 9 October 2022; accepted 12 October 2022; published online 14 October 2022.

*For correspondence (e-mail sandra.smit@wur.nl; mark.aarts@wur.nl).

[†]These authors contributed equally to this work.

[‡]Present address: Faculty of Biology, Ludwig Maximilian University of Munich, Munich, Germany

[†]Present address: Max Planck Institute for Molecular Plant Physiology, Golm, Germany

SUMMARY

Photosynthesis is a key process in sustaining plant and human life. Improving the photosynthetic capacity of agricultural crops is an attractive means to increase their yields. While the core mechanisms of photosynthesis are highly conserved in C₃ plants, these mechanisms are very flexible, allowing considerable diversity in photosynthetic properties. Among this diversity is the maintenance of high photosynthetic light-use efficiency at high irradiance as identified in a small number of exceptional C₃ species. *Hirschfeldia incana*, a member of the Brassicaceae family, is such an exceptional species, and because it is easy to grow, it is an excellent model for studying the genetic and physiological basis of this trait. Here, we present a reference genome of *H. incana* and confirm its high photosynthetic light-use efficiency. While *H. incana* has the highest photosynthetic rates found so far in the Brassicaceae, the light-saturated assimilation rates of closely related *Brassica rapa* and *Brassica nigra* are also high. The *H. incana* genome has extensively diversified from that of *B. rapa* and *B. nigra* through large chromosomal rearrangements, species-specific transposon activity, and differential retention of duplicated genes. Duplicated genes in *H. incana*, *B. rapa*, and *B. nigra* that are involved in photosynthesis and/or photoprotection show a positive correlation between copy number and gene expression, providing leads into the mechanisms underlying the high photosynthetic efficiency of these species. Our work demonstrates that the *H. incana* genome serves as a valuable resource for studying the evolution of high photosynthetic light-use efficiency and enhancing photosynthetic rates in crop species.

Keywords: photosynthesis, Brassicaceae, *Hirschfeldia incana*, model species, high light-use efficiency, high-production agriculture, copy number variation.

INTRODUCTION

Photosynthesis is the biophysical and biochemical process that sustains most life on planet Earth. The most common form of photosynthesis, oxygenic photosynthesis, uses solar energy to convert inorganic carbon dioxide (CO₂) to organic carbon, typically represented as a carbohydrate,

releasing molecular oxygen (O₂) from water in the process. Terrestrial plants provide by far the most conspicuous example of oxygenic photosynthesis (referred to as photosynthesis from now on for brevity) and are responsible for about 50% of the primary production of oxygen in the biosphere, with marine production by eukaryotic algae and

cyanobacteria comprising the other 50%. Agriculture depends on primary production by plants, so expanding our knowledge of photosynthesis is crucial if we are to meet many of the pressing global challenges faced by humankind.

One of these challenges is the need to substantially increase the yield of agricultural crops to meet the increasing demand not only for food and fodder, but also for fibres and similar plant products, and organic precursors for the chemical industry as it transitions away from fossil carbon sources. A major yield-related trait is the conversion efficiency of absorbed solar irradiance to biomass (ϵ_c ; Long et al., 2006), a parameter which is strongly influenced by the light-use efficiency of photosynthesis. As light intensity, or irradiance, increases, the photosynthetic light-use efficiency of leaves and other photosynthetic organs decreases, which leads ultimately to light saturation of photosynthesis (Genty & Harbinson, 1996; Gitelson et al., 2015; Gu et al., 2017; Monneveux et al., 2003; Murchie et al., 1999; Turner et al., 2003). Once light saturation is reached, any additional light will not lead to a further increase in the photosynthetic rate and may even be detrimental to photosynthesis. The threshold for light saturation generally lies far below the maximum level of irradiance experienced in the field or greenhouse (Zhu et al., 2010) and for most C_3 crops this light saturation phenomenon is an aspect of their photosynthesis which remains to be increased in order to increase yield. Improving the photosynthetic light-use efficiency of crop plants thus paves the way towards increasing their ϵ_c and ultimately their yield (Flood et al., 2011; Furbank et al., 2019; Lawson et al., 2012; von Caemmerer & Evans, 2010; Zhu et al., 2010), as recently shown in soybean (*Glycine max*; De Souza et al., 2022).

The means with which to reduce the loss of photosynthetic light-use efficiency in crop plants with increasing irradiance already exists in nature. Most temperate-zone crop species, alongside tropical crops species like rice (*Oryza sativa*), make use of the C_3 photosynthetic pathway, which is the original and ancestral photosynthetic pathway in higher plants, with the alternative CAM and C_4 pathways having evolved as an adaptation to heat and drought, and low CO_2 levels. Due to several issues associated with the C_3 pathway compared to the C_4 pathway, the maximum photosynthesis rates commonly observed among C_3 species are generally lower than those of C_4 ones. Although the core mechanisms of photosynthesis are highly conserved (Leister, 2019; Shi et al., 2005), natural variation in photosynthesis rates has been observed for major crops such as wheat (*Triticum aestivum*; Driever et al., 2014), rice (Gu et al., 2012; Gu et al., 2014), maize (*Zea mays*; Strigens et al., 2013), soybean (Gilbert et al., 2011), and sorghum (*Sorghum bicolor*; Ortiz et al., 2017), as well as for the model species *Arabidopsis thaliana* (van Rooijen et al., 2015; van Rooijen et al., 2017). Much higher photosynthesis rates can be

expected in species that are more ecologically specialised (van Bezouw et al., 2019). Exceptionally high light-use efficiencies (and high assimilation rates) at high irradiance have been found previously in species growing in the Sonoran Desert, such as *Amaranthus palmeri*, *Chylismia claviformis*, *Eremalche rotundifolia*, and *Palafoxia linearis* (Ehleringer, 1985; Werk et al., 1983). Although data collected on these species provided clues about the anatomical and physiological basis of their high photosynthesis rates (Gibson, 1998; Werk et al., 1983), a comprehensive ecophysiological explanation of their phenotypes is still missing.

To understand the physiological and genetic basis of this more efficient photosynthesis at high irradiance, a suitable model species is needed. To date, of the handful of species showing high light-use efficiency that have been described (Ehleringer, 1985; Werk et al., 1983), none would qualify as a model species due to a combination of complex genetics and difficulties in growing in laboratory conditions (e.g. difficult seed germination). Taking inspiration from *A. thaliana*, an attractive model species for high light-use efficiency would need to be easily grown in either regular irradiance (typically up to $600 \mu\text{mol m}^{-2} \text{sec}^{-1}$) and high-light laboratory conditions; have a high-quality reference genome; be a diploid species capable of producing a large number of progeny (hundreds of seeds from a single mother plant) with a short generation time; germinate easily and have easily stored seed; and allow for both inbreeding and outcrossing (Koorneef & Meinke, 2010; Somerville & Koorneef, 2002).

Hirschfeldia incana (L.) Lagr.-Foss. is an excellent candidate that fulfils these requirements. *Hirschfeldia incana* is a thermophilous and nitrophilous annual species native to the Mediterranean basin and the Middle East, but currently widespread in most warm-temperate regions of the world (Siemens, 2011). It is generally self-incompatible and thus allogamous, but a degree of self-compatibility has been observed in natural populations (Lee et al., 2004). Although it makes use of the C_3 pathway, *H. incana* has a very high photosynthesis rate at high irradiance (Canvin et al., 1980), much higher than that of the C_3 crop species wheat (Driever et al., 2014) and rice (Gu et al., 2012), more in the range of C_4 species (Crafts-Brandner & Salvucci, 2002; Leakey et al., 2006). Besides its exceptional physiological properties, *H. incana* is also an attractive model species for practical and genetic reasons. It shows fast and sustained growth in laboratory conditions and is a member of the Brassiceae tribe within the well-studied Brassicaceae family, allowing the use of many genetic and genomic resources developed for the model species *A. thaliana* and its close relatives *Brassica rapa* (Belser et al., 2018; Choi et al., 2007; Kim et al., 2009; The Brassica rapa Genome Sequencing Project Consortium, 2011; Zhang et al., 2018), *Brassica nigra* (Paritosh et al., 2020; Perumal et al., 2020), *Brassica oleracea* (Belser et al., 2018; Liu

et al., 2014; Wang et al., 2011), and *Brassica napus* (Bancroft et al., 2011; Chalhoub et al., 2014). Yet, *H. incana* has received little attention from the research community so far, being recognised mainly as a possible lead (Pb) hyperaccumulator (Auguy et al., 2013; Auguy et al., 2016; Fahr et al., 2015) and for the ecological implications of its occurrence as a weed (Darmency & Fleury, 2000; Lee et al., 2004; Liu et al., 2013; Mira et al., 2019; Sánchez-Yélamo, 2009).

Here we present a genomic assembly and gene set of *H. incana*. We expect these data to lay the foundation for studying photosynthetic light-use efficiency and improving this trait in crop species, through a process of candidate gene identification followed by phenotypic validation using genetic modification and/or gene editing. First, we directly compare the photosynthetic rate of *H. incana* at high irradiance to that of the Brassicaceae species *B. rapa*, *B. nigra*, and *A. thaliana* to affirm its high light-use efficiency. Second, we characterise how the *H. incana* genome differs from that of other members of the Brassicaceae family, specifically focusing on differences in numbers of gene copies. Finally, we report on whether such differences translate to differential expression of genes expected to mediate high light-use efficiency. Our work demonstrates how the genome assembly of *H. incana* serves as a valuable resource to elucidate the genetic basis of high photosynthetic performance and for studying the evolution of this trait in the Brassicaceae family.

RESULTS

Hirschfeldia incana has an exceptionally high rate of photosynthesis

High photosynthesis rates have been reported for *H. incana* in 1980 (Canvin et al., 1980). We performed new measurements in order to compare the performance of *H. incana* with that of close relatives and the well-established model species *A. thaliana* (Figure 1, Table S1). Gross CO₂ assimilation rates differed significantly between these species (Table S2). The two *H. incana* accessions had the highest average gross CO₂ assimilation rates above an irradiance of 550 $\mu\text{mol m}^{-2} \text{sec}^{-1}$ (photosynthetically active radiation), although only 'Burgos' had a statistically significant higher rate than the other species (Table S3). Gross assimilation rates are independent of CO₂ release by mitochondrial respiration and therefore a better indication of photosynthetic capacity than net photosynthesis rates; however, net photosynthesis rates showed a similar trend, but with larger differences between the two *H. incana* genotypes (Figure S1, Table S3), attributed to differences in rates of daytime dark respiration (R_d , Table S4).

A reference genome of *H. incana*

We assembled a scaffold-level reference genome of *H. incana* based on one genotype of the 'Nijmegen' accession.

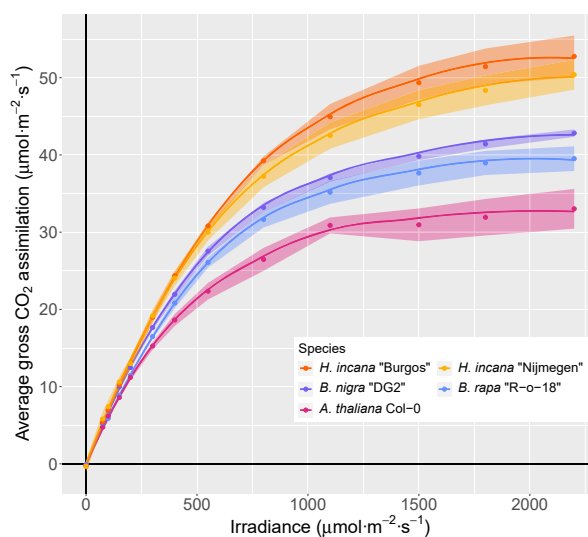


Figure 1. Two *H. incana* genotypes have a higher gross CO₂ assimilation at high irradiance than genotypes of close relatives. Light-response curves for *H. incana*, *B. rapa*, *B. nigra*, and *A. thaliana* accessions adapted to high levels of irradiance. Each point represents the mean gross CO₂ assimilation value of three (*B. rapa*) or four leaves coming from independent plants. Ribbons represent the standard error of the means. The lines indicate trends in gross assimilation for the various species and were obtained via LOESS smoothing.

H. incana is a predominantly self-incompatible species, but 'Nijmegen' produces, nonetheless, some 100 seeds per plant upon self-pollination, which is much more than 'Burgos', so inbreeding is possible. Inbred genotypes are expected to be much more homozygous than open-pollinated genotypes, which is preferred for genome sequencing. Therefore, the 'Nijmegen' accession was inbred for six generations, prior to whole-genome sequencing. Its haploid genome size was estimated to be 487 Mb, based on flow cytometry (Table S5). This estimate is smaller than the previously reported genome size estimates of *B. rapa* (529 Mb) and *B. nigra* (632 Mb) (Johnston et al., 2005). Chromosome counts from root tip squashes showed seven pairs of chromosomes ($2n = 14$) (Figure S2), consistent with previous reports (Anderson & Warwick, 1999; Siemens, 2011).

We generated DNA sequencing data consisting of 56 Gb of PacBio long reads (115-fold genome coverage, based on the genome size estimate), 46 Gb of 10X Genomics synthetic long reads (94-fold coverage, referred to as '10X' from now on for brevity), and 33 Gb of Illumina paired-end short reads (68-fold coverage). In addition, we generated 7.5 Gb of RNA sequencing (RNA-seq) data from leaf tissue for annotation purposes. Summary statistics and accession numbers can be found in Table S6. A k-mer analysis of Illumina data resulted in a haploid genome size estimate of 325 Mb, with a low level of heterozygosity (1.2%).

Using a hybrid assembly strategy, we produced a nuclear genome assembly of 399 Mb of sequence in 384

Table 1 Genomic properties of assemblies generated of *H. incana* 'Nijmegen' (this study), *B. rapa* Chiifu 401-42 (Zhang et al., 2018), and *B. nigra* Ni100 (Perumal et al., 2020)

	<i>H. incana</i>	<i>B. rapa</i>	<i>B. nigra</i>
Technologies	PacBio, 10X, Illumina paired-end	PacBio, BioNano, Hi-C, Illumina mate-pair	Nanopore, Hi-C, genetic mapping
Size (Mb)	398.5	353.1	506
# scaffolds	384	1301	58
N50 (Mb)	5.1	4.4	60.8
Gaps (kb)	0.54	0.40	13
GC content (%)	36.2	36.8	37.0
Complete BUSCO assembly (%)	96.2	97.7	97.0
Complete BUSCO annotation (single/duplicated) (%)	95.1 (80.2/14.9)	97.2 (84.2/13.0)	97.2 (81.9/15.3)
# protein-coding genes	32 313	46 250	59 852
# protein-coding transcripts	38 706	46 250	59 852
Repeat content (%)	49.4	37.5	54.0
Full-length LTR-RTs (%)	25.3	29.2	41.8

scaffolds with an N50 of 5.1 Mb (Table 1, see Table S7 for the full report generated by QUAST; Gurevich et al., 2013). The assembly size is slightly larger than the genome size estimated from Illumina read k-mers (325 Mb), but smaller than the typical overestimate (Sun et al., 2018) based on flow cytometry (487 Mb). Besides the nuclear genome, we assembled the mitochondrial and chloroplast genomes of *H. incana* into single sequences of 253 and 153 kb, and annotated the latter. The chloroplast assembly is typical for a Brassicaceae species, as it is nearly identical to chloroplast assemblies of *A. thaliana*, *B. rapa*, and *B. nigra* in terms of length and number of annotated genes (Table S8), and thus very useful for phylogenetic comparison of *H. incana* with other Brassicaceae.

The nuclear genome assembly is near complete and structurally consistent with the underlying read data of *H. incana* 'Nijmegen' (Table S9). The high mapping rate of Illumina and 10X reads (>93%) suggests completeness, while the lower mapping rate of PacBio reads (81.5%) suggests some misassemblies or missing regions, likely repeats. The high mapping rate of RNA-seq reads (93.6%) also shows the gene space is near complete. We estimated the base-level error rate of the assembly to be 1 per 50 kb at most, based on variant calling using the mapped reads, resulting in 8374 and 4166 homozygous variants from the Illumina and 10X read alignments, respectively.

We have annotated 32 313 gene models and 38 706 transcripts in the *H. incana* assembly (Table 1). This is a conservative annotation, based on filtering 64 546 initial gene models resulting from *ab initio* protein alignment and RNA-seq-based predictions. Our filtering approach is more stringent than those used to generate the *B. rapa* and *B. nigra* annotations, which explains why we report a lower number of genes and transcripts for *H. incana* (Table 1) than for both *Brassica* species. The annotation is expected to cover the large majority of the *H. incana* gene

space. It contains 95.1% of 1440 single-copy orthologues (BUSCOs) conserved in the Embryophyta plant clade, comparable to the percentages found for *B. rapa* and *B. nigra* (both 97.2%) (Table 1). The ratio of single to multiple copies is similar to that of *B. rapa* and *B. nigra* (Table 1), suggesting that the 14.9% of the BUSCOs present in multiple copies are true gene duplications shared by several species of the Brassicaceae tribe. We additionally evaluated the completeness of the annotation by aligning protein sequences of *B. rapa* to the assembly and determining overlap between protein alignments and annotated genes. Out of the 37 387 protein alignments (81.7%), 30 552 corroborate the annotation, as they completely or partially overlap with an annotated protein-coding gene. Of the protein alignments, 2570 (6.9%) completely or partially overlap with an annotated repeat, suggesting that the aligned *B. rapa* proteins correspond to transposable elements (TEs). The remainder of the *B. rapa* proteins completely or partially overlaps with gene models that were filtered (3945, 10.6%) or does not overlap with any annotated element at all (320, 0.9%), indicating a small number of genes that are potentially missing from the annotation. Based on these observations, we conclude that the *H. incana* assembly is mostly contiguous, correct, and complete, making it a solid foundation for comparative analyses with other Brassicaceae.

The genome of *H. incana* extensively diversified from that of *B. rapa* and *B. nigra*

We utilised our assembly to explore the genomic divergence between *H. incana*, *B. rapa*, and *B. nigra*, all members of the same Brassicaceae tribe. A substantial degree of divergence is expected between the three species due to different processes of post-polyploid diploidization, i.e. the process in which polyploid genomes get extensively rearranged as they return to a diploid state (Mandáková &

Lysak, 2018), following the ancient two-step genome triplification event shared by all Brassiceae (He et al., 2021; Lysak et al., 2005; The Brassica rapa Genome Sequencing Project Consortium, 2011). Part of this divergence may have facilitated the evolution of the exceptionally high rate of photosynthesis at high irradiance in *H. incana*.

We first assessed the phylogenetic relationship between *H. incana*, *B. rapa*, and *B. nigra* by constructing phylogenetic trees based on homologous nuclear and chloroplast genes, using *A. thaliana* as the outgroup. Both trees are congruent with each other and suggest that *H. incana* is more closely related to *B. nigra* than *B. rapa* (Figure 2a,b). This is corroborated by the median rate of synonymous substitutions between the syntenic orthologues (K_s) of the three species, which correspond to speciation events with an estimated time of 10.4 (*H. incana* and *B. nigra*) and 11.6 (*H. incana* and *B. rapa*) million years ago (mya) (Figure 2c) as obtained by dividing the median K_s of each curve by the rate of 8.22×10^{-9} synonymous substitutions per year established for Brassicaceae species (Beilstein et al., 2010). Our results are consistent with a previous phylogenetic analysis based on four intergenic chloroplast regions (Arias & Pires, 2012), but slightly differ from a more recently constructed phylogeny of the Brassicaceae, which was only based on 113 nuclear genes (Huang et al., 2016), while we consider many more.

We determined rearrangements between the genomes of *H. incana* and *B. rapa* and between those of *H. incana* and *B. nigra* by comparing the order of syntenic orthologues between their assemblies. On a small scale, most genomic regions of *H. incana* are syntenic (not rearranged) with *B. rapa* and *B. nigra*, as 77.7% and 81.0% of the genes of *H. incana* could be clustered in collinear blocks containing a minimum of four orthologous pairs of *H. incana* and *B. rapa* and of *H. incana* and *B. nigra*, respectively. Gene order is less conserved when comparing larger blocks, indicating several rearrangements between the 20 largest

scaffolds of *H. incana* (covering 43.6% of the assembly) and the chromosomes of the other two species (Figures 3a and S3). For example, the two largest scaffolds of the *H. incana* assembly both contain inversions and/or translocations relative to their homologous chromosomes in *B. rapa* and *B. nigra*. A similar pattern of rearrangements of small collinear blocks was observed between the genomes of *B. rapa* and *B. nigra* in previous work (He et al., 2021).

We further examined genomic differentiation between the three species by comparing their TE content. Of the assembly of *H. incana*, 49.4% consists of repetitive elements (Table 1), of which most are long terminal repeat retrotransposons (LTR-RTs) (25.3% of the genome). These numbers are consistent with previous work that investigated the repeat content of the *H. incana* genome using genome skimming, and which reported a repeat content of 46.5% and an LTR-RT content of 31.6% (Beric et al., 2021). We specifically focused our analyses on LTR-RTs, as LTR-RT expansion and contraction has been previously identified as a major driver of genomic differentiation between Brassicaceae (C. Xu et al., 2018), even between different ecotypes of the same species (Cai et al., 2020). The composition of LTR-RTs in the *H. incana* assembly differs from that of the *B. rapa* and *B. nigra* assembly, as the majority of LTR-RTs consist of Gypsy elements in *H. incana*, consistent with earlier work (Beric et al., 2021), while Copia retrotransposons form the majority of LTR-RTs in the others (Figure 3b). Furthermore, the estimated insertion times of LTR-RTs vary between the three assemblies, as Gypsy and Copia elements in *H. incana* and *B. rapa* are predicted to have proliferated recently (<1 mya) (Figure 3c,d), while Gypsy elements in *B. nigra* show a more varied distribution of insertion times (Figure 3c). A possible explanation of this shift could be that the *B. nigra* assembly was generated using longer reads than those used for the assemblies of *H. incana* and *B. rapa*, enabling it to capture a larger proportion of the centromeric regions, but we found no

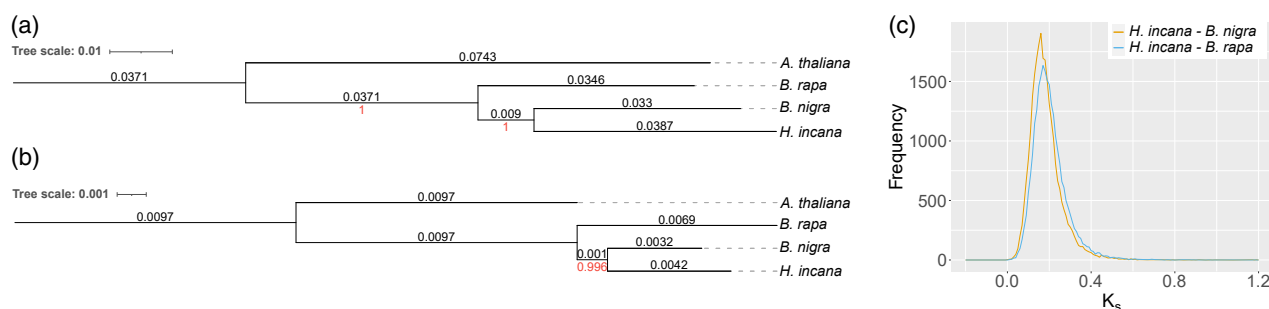


Figure 2. *Hirschfeldia incana* is more closely related to *B. nigra* than *B. rapa*. (a,b) Phylogenetic trees of *H. incana*, *B. rapa*, *B. nigra*, and *A. thaliana* (outgroup), based on nuclear (a) and chloroplast (b) genes. Branch lengths (black) and bootstrap values (red) are displayed above and below each branch, respectively. (c) Distributions of the rates of synonymous substitutions between 25 127 and 26 137 orthologous gene pairs of *H. incana* and *B. rapa* and between *H. incana* and *B. nigra*, respectively. Both distributions show a single peak corresponding to speciation events with an estimated time of 11.6 (*H. incana* and *B. rapa*) and 10.4 (*H. incana* and *B. nigra*) mya.

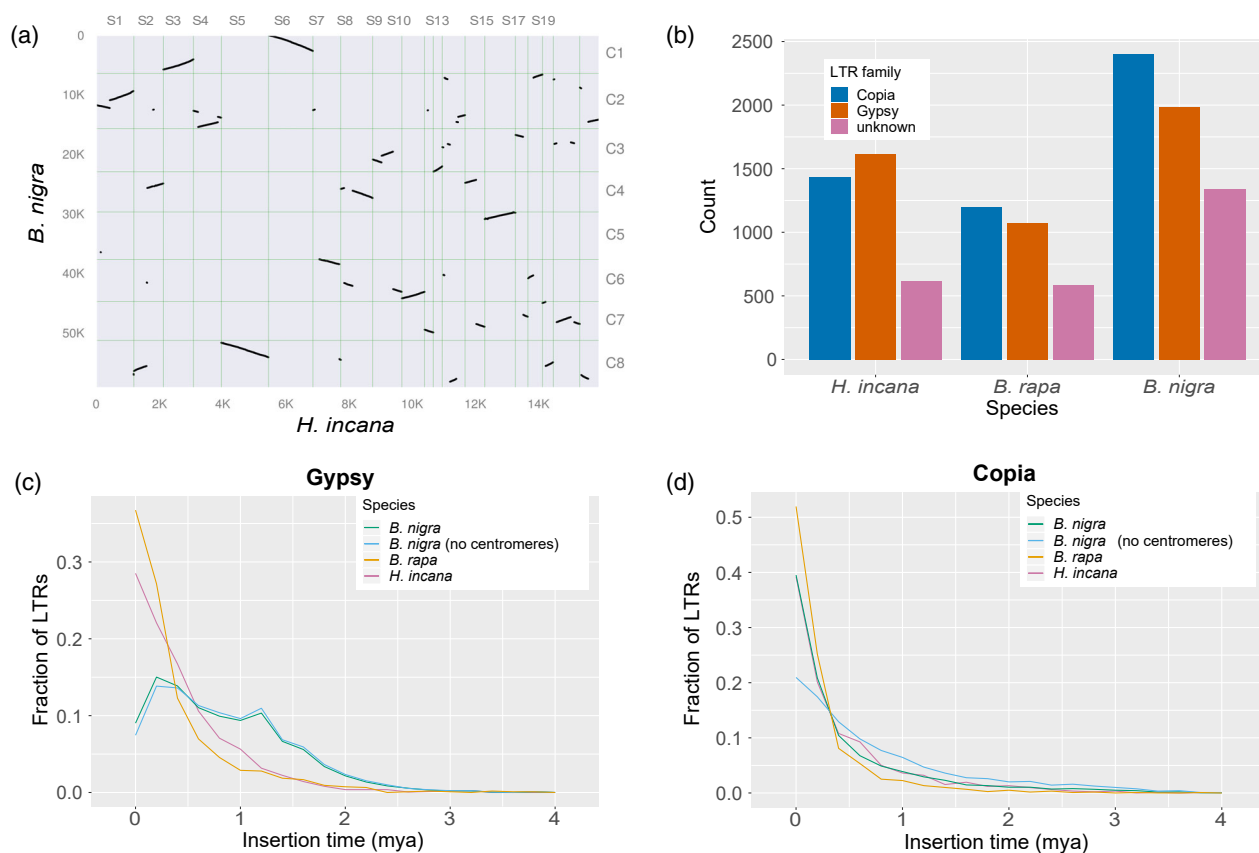


Figure 3. The genome of *H. incana* extensively diversified from that of *B. rapa* and *B. nigra*. (a) Orthologous syntenic blocks between the genomes of *H. incana* and *B. nigra*. Dots indicate pairs of syntenic orthologs that are found in the same order in both genomes according to sequence positions. Only the 20 largest scaffolds of *H. incana* (43.6% of the assembly) are shown for clarity. Axis labels correspond to the total number of genes annotated on the sequences (left and bottom) and identifiers of the scaffolds (top) or chromosomes (right). A dot plot visualising orthologous syntenic blocks between *H. incana* and *B. rapa*, showing similar patterns, is found in Figure S3. (b) Frequency distribution of long terminal repeat retrotransposon (LTR-RT) families. LTR-RTs are classified as unknown if they contained elements of both Gypsy and Copia sequences and could thus not be reliably assigned to either of these families. (c) Frequency polygon (bin width = 0.2 mya) of the insertion times of Gypsy elements. (d) Frequency polygon (bin width = 0.2 mya) of insertion times of Copia elements.

evidence that this introduced a bias towards longer insertion times of Gypsy elements (Figure 3c).

Taken together, the breakdown of genomic synteny and divergence of LTR-RT content indicates that the genome of *H. incana* extensively diversified from that of *B. rapa* and *B. nigra* following their shared genome triplication event.

Gene copy number expansion may contribute to high photosynthetic rates

Genomic differentiation can result in species-specific gains and losses of genes, and these may explain the differences in photosynthetic light-use efficiency between *H. incana*, *B. rapa*, and *B. nigra*. Given that the three species all share the same ancient genome triplication event (He et al., 2021; Schranz et al., 2006), it is reasonable to assume that most differences originated through differential retention of duplicated genes, particularly those located in genomic blocks showing evidence of extensive fractionation since that event (He et al., 2021). We investigated gene copy number variation between the three species by clustering

their annotated protein-coding genes with those of five other Brassicaceae species within (*Raphanus raphanistrum* and *Raphanus sativus*) and outside (*Aethionema arabicum*, *A. thaliana*, and *Sisymbrium irio*) the Brassicaceae tribe into homology groups. The inclusion of *A. thaliana* allowed us to use its extensive genomic resources to functionally annotate the genes of the other species. The other four species were included to put the analysis into a broader phylogenetic context. *Aethionema arabicum* is part of the Aethionema tribe, which diverged from the core group of the Brassicaceae family, thus allowing us to identify highly conserved genes. *Sisymbrium irio* is part of a different tribe than *A. thaliana* (Sisymbrieae) that is more closely related to the Brassicaceae tribe (Huang et al., 2016), but did not undergo the ancient genome triplication. *Raphanus raphanistrum* and *R. sativus* are part of the *Raphanus* genus within the Brassicaceae tribe and thus represent another set of species that underwent the genome triplication event shared by the whole tribe.

Our analysis resulted in 20 331 groups containing at least one *H. incana* gene (Table S10). The composition of the homology groups agrees with the currently established phylogeny of the Brassicaceae (Huang et al., 2016) as groups containing *H. incana* genes share the least number of genes with *A. arabicum* (58.2%) and the greatest number of genes with species that are part of the Brassicaceae tribe (86.3–95.6%). *Hirschfeldia incana* has a low fraction of species-specific homology groups (3.4%) compared to the seven other species, which can be attributed to the stringent filtering of the predicted gene models.

We focused on a subset of 15 097 groups containing at least one gene of *A. thaliana* and one of *H. incana*, as these could be extensively annotated through the transfer of Gene Ontology (GO) terms from *A. thaliana* genes to their respective groups. Consistent with the expectation that most genes quickly return to single-copy status following a whole-genome duplication event (Z. Li et al., 2016), 70.2% of these groups contain a single gene of both *A. thaliana* and *H. incana*. Focusing on groups containing *A. thaliana* genes involved in photosynthesis (260 in total, Table S11), most contain a higher number of genes of *H. incana*, *B. rapa*, and *B. nigra* compared to *A. thaliana* (Figure 4), consistent with the relatively higher photosynthetic light-use efficiency of *H. incana*, *B. rapa*, and *B. nigra* (Figure 1). That *H. incana* should have the highest efficiency of all the four species is not apparent from the gene copy number data because for most groups of genes, *H. incana* contains the same or a lower number of copies relative to *B. rapa* and *B. nigra*. This is not a result of our conservative filtering approach, as we explicitly retained putative photosynthesis-related genes during our filtering procedure (Methods S1). Besides photosynthesis-related genes, we also analysed copy numbers of a more general set. We found that 4901 homology groups contain genes of which the copy number in *H. incana* is higher than in *A. thaliana* and equal to or higher than those of *B. rapa* and *B. nigra* (Table S12). We estimate that 74.5% of the duplicated gene pairs in *H. incana* (16 788 of the 22 535 analysed pairs) were duplicated through whole-genome duplication, 1.8% through tandem duplication, and the remaining 23.6% through another mode of duplication. Given that the increased photosynthetic light-use efficiency of *H. incana* relative to *A. thaliana*, *B. rapa*, and *B. nigra* is particularly pronounced at high levels of irradiance (Figure 1), genes annotated with GO terms associated with photosynthesis and/or photoprotection are of particular interest. The 4901 homology groups contain ample examples of such genes (Table S12), although the groups are not significantly enriched for any GO term specifically linked to photosynthesis and/or photoprotection (Table S13).

As gene copy number variation can considerably affect expression levels (Zmieńko et al., 2014), we hypothesised that retained copy number expansions of

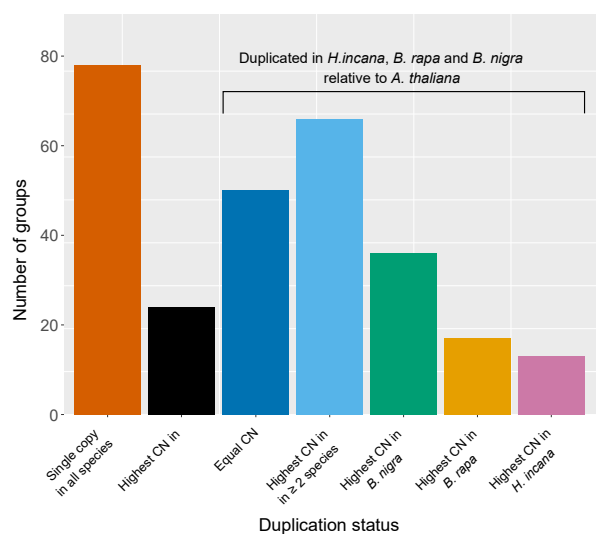


Figure 4. *Hirschfeldia incana* retained fewer duplicated copies of photosynthesis-associated genes than *B. rapa* and *B. nigra*. Bars show counts of homology groups containing genes associated with photosynthesis with different distributions of copy numbers (CNs) in the four species (260 groups in total). For groups that contain a higher number of copies in *H. incana*, *B. rapa*, and *B. nigra* than in *A. thaliana*, it has been indicated whether the same number of copies is found in all three species (equal CN), whether there are two or more species that contain a higher number of copies than the other(s) (highest CN in at least two species), or whether there is a single species containing the highest number of copies.

photosynthesis- and photoprotection-associated genes in *H. incana*, *B. rapa*, and *B. nigra* may contribute to the high photosynthetic capacities of these species (Figure 1). We therefore measured gene expression levels of nine genes for which there is inter-species copy number variation in two contrasting light conditions ($200 \mu\text{mol m}^{-2} \text{sec}^{-1}$ and $1800 \mu\text{mol m}^{-2} \text{sec}^{-1}$), selecting genes with a function related to photosynthesis and/or photoprotection (Table 2). *Arabidopsis thaliana*, the species with the lowest photosynthesis rates measured in this study, contains a single copy of each of the tested genes. For six genes, we observed a statistically significant positive correlation between gene expression level and gene copy number (Figure 5), with species showing higher or equal expression with an increasing number of copies (Figure S4). No such correlation was observed for the remaining three genes.

To test if the observed differences in gene expression are due to photosynthesis-related genes being more frequently more highly expressed in general in *B. rapa*, *B. nigra*, and *H. incana* compared to *A. thaliana*, we included nine additional genes in our experiment that are present in a single copy in all four species and involved in similar processes as the multi-copy genes. Although we find species-specific differences in expression for this set of genes, no consistently higher gene expression levels are

Table 2 Genes with inter-specific copy number variation of which expression was measured. All genes are annotated to a function in photosynthesis and/or photoprotection

Gene name	Full gene name	Copy number			
		<i>A. thaliana</i>	<i>B. rapa</i>	<i>B. nigra</i>	<i>H. incana</i>
LHCA6	PHOTOSYSTEM I LIGHT HARVESTING COMPLEX GENE 6 (Peng et al., 2009)	1	1	1	3
ELIP1	EARLY LIGHT INDUCED PROTEIN (Hutin et al., 2003)	1	3	3	3
SIGE/SIG5	SIGMA FACTOR 5 (Tsunoyama et al., 2004)	1	2	2	2
SIGD/SIG4	SIGMA FACTOR 4 (Favory et al., 2005)	1	2	2	3
BBX21	B-BOX DOMAIN PROTEIN 21 (Crocco et al., 2018)	1	3	3	3
PETC	PHOTOSYNTHETIC ELECTRON TRANSFER C (Maiwald et al., 2003)	1	2	2	2
ABC1K3	ABC1-LIKE KINASE 3 (Martinis et al., 2013)	1	1	1	2
OHP2	ONE-HELIX PROTEIN 2 (Y. Li et al., 2019)	1	2	2	3
CYFBP	CYTOSOLIC FRUCTOSE-1,6-BISPHOSPHATASE (S.-K. Lee et al., 2008)	1	2	2	2

found in *B. rapa*, *B. nigra*, and *H. incana* compared to *A. thaliana* (Figure S5). Overall, our analyses suggest that the increased copy numbers of photosynthesis- and photoprotection-associated genes in *H. incana*, *B. rapa*, and *B. nigra* relative to *A. thaliana* may contribute to their high photosynthetic efficiency, although this effect appears to not be specific to a particular species or level of irradiance.

DISCUSSION

In this study, we generated a reference genome of *H. incana* to establish this species as a model for exceptional photosynthetic light-use efficiency at high irradiance. We find substantial differences in light-use efficiency, genomic structure, and gene content between *H. incana* and its close relatives. We discuss these results in terms of how they contributed to the evolution of the remarkable phenotype of *H. incana*.

Our results show an even higher photosynthetic light-use efficiency at high irradiance than previously reported for *H. incana* (Canvin et al., 1980), with photosynthesis rates varying marginally between both accessions. Examination of a wider set of *H. incana* accessions may identify genotypes with larger differences in photosynthesis rates, which would allow a quantitative genetic approach to identify alleles conferring high photosynthesis rates. Our measurements imply that the photosynthesis rates of this C_3 species are higher than those of the C_4 crop maize (Crafts-Brandner & Salvucci, 2002; Leakey et al., 2006) and almost

two times higher than those typically reported from key cereal crop species with C_3 photosynthetic metabolism, such as wheat (Driever et al., 2014) and rice (Gu et al., 2012). Furthermore, these rates are higher than those of the closely related Brassicaceae species *B. rapa* and *B. nigra* and the more distantly related *A. thaliana*. The photosynthesis rates we measured in *B. rapa* are also higher than previously reported (Pleban et al., 2018; Taylor et al., 2020). Although the rates presented in this study were obtained from plants grown in controlled, favourable conditions and thus could be an overestimation of rates in natural environments, the magnitude of the differences suggests that the *H. incana* genome holds essential information for the improvement of photosynthetic light-use efficiency in crops.

The reference genome of *H. incana* generated in this study provides the means to elucidate the genetic basis of this plant's exceptional rate of photosynthesis and how it evolved in this species. We estimate that *H. incana* diverged 11.6 and 10.4 mya from *B. rapa* and *B. nigra*, respectively, consistent with an earlier study that used a smaller set of nuclear genes (Huang et al., 2016). These time points are close to the reported time at which *B. rapa* and *B. nigra* (Perumal et al., 2020) diverged from each other (11.5 mya) and the time at which the whole Brassicaceae family underwent a rapid radiation event (Franzke et al., 2009). This event may have been mediated by the expansion of grass-dominated ecosystems in the region inhabited by Brassicaceae family members at that time,

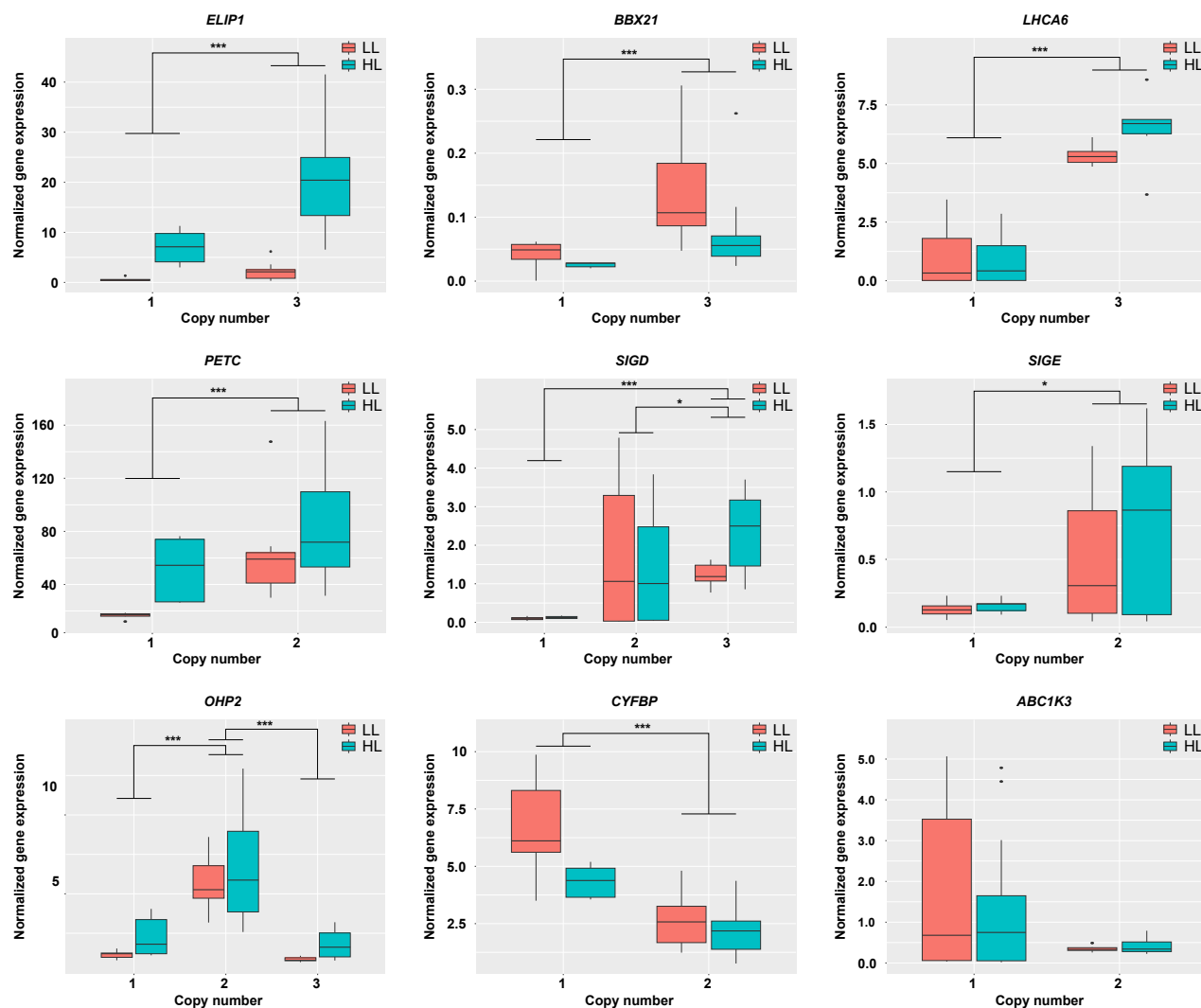


Figure 5. Copy numbers of photosynthesis- and photoprotection-associated genes correlate with expression level. Boxplots depict gene expression levels of *A. thaliana*, *B. rapa*, *B. nigra*, and *H. incana* grown in $200 \mu\text{mol m}^{-2} \text{sec}^{-1}$ and $1500 \mu\text{mol m}^{-2} \text{sec}^{-1}$. Gene expression levels were normalised against *H. incana* grown at $200 \mu\text{mol m}^{-2} \text{sec}^{-1}$ and subsequently grouped per gene copy number. Titles of graphs indicate gene names based on the *A. thaliana* gene nomenclature. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

which created new open habitats that favoured rapid diversification (Franzke et al., 2009). This expansion of grasslands is thought to have been driven by decreasing atmospheric CO_2 levels and increasing aridity, which favoured the displacement of the then dominant C_3 plants by C_4 grasses (Edwards et al., 2010). We argue that climatic changes also drove the evolution of the high photosynthetic rates observed in *H. incana*; grassland, i.e. non-forested ecosystems, may have provided the ephemeral niches with high irradiances that favoured the evolution of high photosynthetic rates. Species with high photosynthetic rates are currently found in Mediterranean and desert ecosystems (Ehleringer, 1985; Werk et al., 1983). The evolution of high rates of C_3 photosynthesis could therefore have paralleled the expansion of the C_4

photosynthesis pathway as an adaptation to low CO_2 levels and drought.

Our analyses suggest that the genome of *H. incana* extensively differentiated from the genomes of *B. rapa* and *B. nigra* since their time of divergence, through large genomic arrangements and differences in LTR-RT content. Previous analyses of natural *A. thaliana* accessions indicated that specific LTR-RT families show increased rates of proliferation in response to particular types of environmental stress (Baduel et al., 2021), which may explain the species-specific amplification of Gypsy elements that we observed in *H. incana*. Such elements may have been retained because this particular LTR-RT family generally inserts outside of exons (Baduel et al., 2021). We hypothesise that the differences in LTR-RT content between *H. incana*, *B. rapa*,

and *B. nigra* were caused in part by Gypsy elements being less efficiently purged from the genome of *B. nigra* than from the others. An increased rate of LTR-RT removal, based on the ratio of solo LTRs to intact LTR-RTs, has also been observed in *B. rapa* relative to *B. oleracea* and it was speculated that this is caused by the increased rate of genetic recombination in the former (Zhao et al., 2013). Given that a similar negative correlation between local recombination rate and LTR-RT content was found in rice (Tian et al., 2009), soybean (Du et al., 2012), and eukaryotes in general (Kent et al., 2017), the differences in predicted insertion times of Gypsy elements in *H. incana*, *B. rapa*, and *B. nigra* observed in this study may thus reflect different rates of genetic recombination in the three species. While it has been suggested that changes in recombination rate can be adaptive, there is little empirical evidence that supports this (Ritz et al., 2017). It would therefore be interesting to directly measure genome-wide rates of recombination of *H. incana*, *B. rapa*, and *B. nigra* and explore whether these are correlated with their rates of photosynthesis.

Further comparative analyses between the genomes of *H. incana*, *B. rapa*, *B. nigra*, and *A. thaliana* revealed numerous species-specific gains and losses of genes. For dosage-sensitive genes, such as those involved in transcriptional regulation, differences may not necessarily reflect adaptive selection. This category of genes was found to be consistently retained in multiple copies following polyploidy events across the Brassicaceae (Mandáková et al., 2017) and a wide group of angiosperms (Z. Li et al., 2016), which is hypothesised to be due to dosage constraints (Edger & Pires, 2009). Differences in copy number of such genes may thus reflect different rates of relaxation of dosage balance constraints and subsequent loss of duplicates through time, which is a neutral process.

On the other hand, there is reason to believe that gene duplications contributed to the evolution of the high light-use efficiency of *H. incana*. Gene duplications have been identified as important drivers of plant evolution and differences in gene copy number between species are often enriched for adaptive evolutionary traits (Dassanayake et al., 2011; Oh et al., 2013; Rizzon et al., 2006; Suryawanshi et al., 2016). Moreover, reverse transcriptase quantitative PCR (RT-qPCR) analysis of nine duplicated genes associated with photosynthesis and/or photoprotection showed that the expression levels of six of them correlate with gene copy number. In contrast, nine photosynthetic genes present in a single copy in all species did not show significantly increased expression in *H. incana*, *B. nigra*, and *B. rapa* compared to *A. thaliana*, indicating that photosynthetic genes are not overexpressed in the former three species in general. This supports a putative role for gene duplications in mediating the high light-use efficiency achieved by *H. incana*, *B. nigra*, and *B. rapa*.

The most striking genes of which copy number correlated with gene expression are *LHCA6* and *ELIP1*, involved in response to high light and having the highest expression in *H. incana* growing under high light (Figure S4). *LHCA6* encodes a light-harvesting complex I (LHCI) protein of photosystem I (PSI) that together with LHCA5 is required to form a full-size NAD(P)H dehydrogenase (NDH)-PSI supercomplex (Peng et al., 2009). Higher expression of *LHCA6* might help the formation of the NDH-PSI complex, thought to help stabilise NDH under high-irradiance conditions. In turn, NDH has proposed roles supporting the Calvin-Benson cycle's activity (Harbinson et al., 2022) and photoprotection by preventing overreduction at high light intensities (Munekage et al., 2004). *ELIP1* encodes a protein with a proposed role in photoprotection and is associated with high-light stress (Heddad et al., 2006; Norén et al., 2003; Youssef et al., 2010). Increased expression of this gene is expected to make the photosynthetic apparatus of *H. incana* more resistant to photoinhibition at high levels of irradiance. While the *H. incana* genome harboured the highest number of copies of *LHCA6* compared to the genomes of *A. thaliana*, *B. rapa*, and *B. nigra*, this is not the case for *ELIP1*, for which *H. incana*, *B. nigra*, and *B. rapa* all have three copies as opposed to the single copy of *A. thaliana*. Therefore, although we can propose a role for gene duplications in achieving higher light-use efficiency, the exact nature of this role still remains unclear as it appears to not be completely dependent on species or light treatment.

While our gene expression analysis provides several promising leads, it only offers a glimpse of what may contribute to the high photosynthetic light-use efficiency of *H. incana*. Besides the nine genes included in this analysis, we identified many more genes with a high copy number in *H. incana* that warrant further investigations on a whole-transcriptome level. Such investigations should not be limited to core photosynthetic genes, as *H. incana* can only attain high photosynthetic light-use efficiency through changes in many other traits that are outside the chloroplast, such as leaf architecture affecting mesophyll conductance to CO₂, the synthesis of carbohydrates in the cytosol, the transport of carbohydrates from the leaf, the uptake from the soil and the supply of nitrogen and other minerals to the leaf, the abundance and distribution of different leaf pigments, and (photo)respiration. Nor should they include duplicated genes only, as it is striking that *H. incana* shows a higher light-use efficiency than *B. rapa* and *B. nigra*, though it contains fewer photosynthesis-related genes than the latter two species. This points towards alternative scenarios in which adaptation of *H. incana* photosynthesis to high levels of irradiance occurred through regulation of expression of one copy of the photosynthesis-related genes, which relaxed selection on duplicate retention or even encouraged loss of duplicate copies, or through other traits, as described above.

To elucidate the exact genetic mechanisms underlying the high light-use efficiency of *H. incana*, a natural follow-up to this study is to perform comparative transcriptomic analyses of leaves of *H. incana*, *B. rapa*, and *B. nigra* under a range of different levels of irradiance and at different developmental stages. Genes that show copy number variation and are differentially expressed between *H. incana* and the latter two species, such as *LHCA6*, would then be prime candidates to further test for potential causality through e.g. knock-out mutant analysis. As previous work has shown that it is possible to cross distantly related Brassicaceae species (Katche et al., 2019), a useful approach to further pinpoint the causal genes is to establish a genetic mapping population between *H. incana* and a Brassicaceae species with regular light-use efficiency and perform quantitative trait locus analyses of photosynthetic traits segregating within the population. It would also be useful to expand comparative genome and transcriptome analyses to plant species outside of the Brassicaceae family that show high photosynthetic light-use efficiency, such as the aforementioned *A. palmeri*, *C. claviformis*, *E. rotundifolia*, and *P. linearis*. Such expanded analyses could be informative, for instance to investigate amino acid substitutions or lateral gene transfer specific to species with high photosynthetic light-use efficiency. Furthermore, transcriptome data may indicate genes showing differences in expression between such species and those that are less efficient, providing further insight into which genes contribute to the evolution of this trait.

CONCLUSIONS

Hirschfeldia incana has an exceptional rate of photosynthesis at high irradiance. We generated a near-complete reference genome of this species and found evidence suggesting that its exceptional rate evolved through differential retention of duplicated genes. Taken together, our work provides several promising leads that may explain the high photosynthetic light-use efficiency of *H. incana*, and we expect the reference genome generated in this study to be a valuable resource for improving this efficiency in crop cultivars.

MATERIALS AND METHODS

Plant material

Hirschfeldia incana accessions 'Nijmegen' and 'Burgos' were used. 'Nijmegen' is an inbred line (more than six rounds of inbreeding) originally collected in Nijmegen, the Netherlands. Seeds of 'Burgos' were originally collected near Burgos, Spain. Furthermore, *B. nigra* accession 'DG2', sampled from a natural population near Wageningen, the Netherlands, the *B. rapa* inbred line 'R-o-18' (Bagheri et al., 2012; Stephenson et al., 2010), and *A. thaliana* accession Col-0 were used.

Measurements of photosynthesis rates

Seeds of *H. incana* 'Nijmegen', *H. incana* 'Burgos', *B. rapa* 'R-o-18', *B. nigra* 'DG2', and *A. thaliana* Col-0 were sown in 3-L pots

filled with a peat-based potting mixture. Plants were grown in a climate chamber with a 12/12-h photoperiod at day and night temperatures of 23°C and 20°C, respectively. Relative humidity and CO₂ levels were set at 70% and 400 ppm, respectively. The chamber was equipped with high-output LED light modules (VYPR2p, Fluence by OSRAM). Plants were watered daily with a custom nutrient solution (0.6 mM NH₄⁺, 3.6 mM K⁺, 2 mM Ca²⁺, 0.91 mM Mg²⁺, 6.2 mM NO₃⁻, 1.66 mM SO₄²⁻, 0.5 mM P, 35 μM Fe³⁺, 8 μM Mn²⁺, 5 μM Zn²⁺, 20 μM B, 0.5 μM Cu²⁺, 0.5 μM Mo⁴⁺). The seeds were germinated at an irradiance of 300 μmol m⁻² sec⁻¹, and the same irradiance was maintained to let seedlings establish. On days 14, 21, and 25 after sowing, the irradiance was raised to 600, 1200, and 1800 μmol m⁻² sec⁻¹, respectively.

The photosynthetic metabolism of young, fully expanded leaves developed under 1800 μmol m⁻² sec⁻¹ of light was measured with a LI-COR 6400XT portable gas exchange system (LI-COR Biosciences) equipped with a 2-cm² fluorescence chamber head. 'Rapidly' descending light-response curves were measured between 30 and 35 days after sowing to accommodate differences in growth rates of the different species on one leaf from four *H. incana* 'Nijmegen', *H. incana* 'Burgos', *B. nigra* 'DG2', and *A. thaliana* Col-0 plants and three *B. rapa* 'R-o-18' plants. The net assimilation rates of the plants were measured at 13 different levels of irradiance ranging from 0 to 2200 μmol m⁻² sec⁻¹. During measurements, leaf temperature was kept constant at 25°C and the reference CO₂ concentration was kept at 400 μmol mol⁻¹. Water in the reference air flux was regulated in order to achieve vapour pressure deficit values between 0.8 and 1.2 kPa.

Light-response curve parameters (A_{max} : net CO₂ assimilation at saturating irradiance; ϕ : apparent quantum yield of CO₂ assimilation; R_d : daytime dark respiration rate; and θ : curve convexity) were estimated for each species through non-linear least squares regression of a non-rectangular hyperbola (Marshall & Biscoe, 1980) with the R package 'photosynthesis' (version 2.0.0) (Stinziano et al., 2020). An indication of gross assimilation rates for each species was subsequently generated by adding the daytime dark respiration rate (R_d) estimated for each species to the species' net assimilation rates.

Differences in net and gross assimilation rates were tested at each light level of the light-response curve with one-way analysis of variance (ANOVA) on the 'genotype' experimental factor. Pairwise comparisons between the assimilation rates of the different genotypes at each light level were subsequently performed and tested with the Tukey-Kramer extension of Tukey's range test. The *P*-value threshold for statistical significance was set at $\alpha = 0.05$.

Flow cytometry

Leaf samples of the *H. incana* genotypes 'Burgos' and 'Nijmegen' and *A. thaliana* Col-0 were analysed for nuclear DNA content by flow cytometry (Plant Cytometry Services B.V., Didam, the Netherlands). Seven, three, and five biological replicates were measured over separate rounds of analysis for *H. incana* 'Nijmegen', *H. incana* 'Burgos', and *A. thaliana* Col-0, respectively. Nuclei were extracted from leaf samples following the method by Arumuganathan & Earle, 1991, and stained with 4',6-diamidino-2-phenylindole (DAPI). The DNA content of nuclei relative to that of the reference species *Monstera deliciosa* was determined on a CyFlow Ploidy Analyser machine (Sysmex Corporation, Kobe, Japan). A haploid flow cytometry estimate of 157 Mb was used for *A. thaliana*, resulting from comparisons of nuclear DNA content of this species and other model organisms (Bennett et al., 2003). Haploid genome size estimates for the *H. incana*

genotypes were obtained by multiplying the *H. incana*-to-*M. deliciosa* ratio by the haploid *A. thaliana* estimate and dividing this product by the average *A. thaliana*-to-*M. deliciosa* ratio.

Chromosome counting

Root tips (approximately 1 cm long) were collected from young, fast-growing rootlets of multiple *H. incana* 'Nijmegen' plants and pre-treated for 3 h at room temperature with a 0.2 mM 8-hydroxyquinoline solution. After pre-treatment, the 8-hydroxyquinoline solution was replaced with freshly prepared Carnoy fixative (1:3 [v/v] acetic acid:ethanol solution) and maintained at room temperature for half a day. Root tips were then rinsed with 70% ethanol three times to remove remaining fixative and stored in 70% ethanol at 4°C until further use. Prior to slide preparation, root tips were rinsed twice in Milli-Q (MQ) water before adding 1:1 solution of a pectolytic enzymatic digestion solution (1% Cellulase from *Trichoderma*, 1% Cytohellicase from *Helix Promatia*, 1% Pectolyase from *Aspergillus japonicus*) and 10 mM citric buffer. After incubation at 37°C for 1 h, the enzymatic digestion solution was replaced by MQ water. The digested root tips were spread in 45% acetic acid over microscopy slides on a hot plate set at 45°C, and cells were fixed with freshly prepared Carnoy fixative, dried, and stained with DAPI dissolved in Vectashield mounting medium (Vector Laboratories Inc., Burlingame, US). Slides were imaged with an Axio Imager.Z2 fluorescence optical microscope coupled with an Axiocam 506 microscope camera (Carl Zeiss AG, Oberkochen, Germany) at 63× magnification. Chromosome numbers were counted in metaphase mitotic cells and averaged to obtain the reported number.

DNA and RNA isolation

Genomic DNA was extracted from *H. incana* 'Nijmegen' samples using a protocol modified from Chang et al., 1993. The modifications consisted of adding 300 µL β-mercaptoethanol to the extraction buffer just before use. We added 0.7% isopropanol to the supernatant instead of 10 mM LiCl and then divided the total volume into 1-ml aliquots for subsequent extractions. The pellet was dissolved in 500 µL of SSTE which was pre-heated to 50°C before use. The final pellets were dissolved in 50 µL MQ water and then pooled at the end of the extraction process. DNA used for Illumina and 10X Genomics sequencing was extracted from flower material, while leaf material was used for PacBio sequencing, all originating from the same plant.

Total RNA was extracted from leaf material of *H. incana* 'Nijmegen' from a different plant than the one used for the DNA isolations with the Direct-zol RNA mini-prep kit (Zymo Research, Irvine, CA, USA) according to the company's instructions and then subjected to DNase (Promega Corporation, Madison, WI, USA) treatment at 37°C for 1 h.

Generation of sequencing data

Sequencing of total-cellular DNA of *H. incana* 'Nijmegen' was performed by GenomeScan B.V., Leiden. A total of seven SMRT cells were used for sequencing on the Pacific Biosciences Sequel platform. Short-read Illumina and 10X Genomics libraries with an insert size of approximately 500–700 bp were prepared with the NEBNext Ultra DNA Library Prep kit for Illumina and the 10X Genomics Chromium™ Genome v1 kit, respectively. These libraries were sequenced using an Illumina X10 platform (2 × 151 bp). RNA paired-end sequencing libraries with an average insert size of 254 bp were prepared using the Illumina TruSeq RNA sample prep kit with polyA mRNA selection and sequenced using an Illumina HiSeq 2500 platform (2 × 125 bp).

k-Mer analysis

A histogram of k-mer frequencies of Illumina reads predicted to be of nuclear origin (Methods S1) was generated using Jellyfish (v2.2.6) (Marçais & Kingsford, 2011) using a k-mer length of 21. The resulting histogram was provided as input to GenomeScope (v1.0.0) (Vurture et al., 2017) to estimate genome size and heterozygosity.

Genomic assembly and annotation

The genome assembly and annotation process is more extensively described in Methods S1. In short, we generated an initial assembly based on the PacBio data only with Canu (Koren et al., 2017) and used it to bin the PacBio, 10X, and Illumina reads according to whether they originated from nuclear, organellar, or contaminant DNA. The bins were used to separately assemble the nuclear and organellar genomes, yielding a nuclear assembly consisting of hundreds of contigs and mitochondrial and chloroplast assemblies that were both represented by a single sequence. Nuclear contigs representing alternative haplotypes were removed using purge_dups (Guan et al., 2020), after which ARKS (Coombe et al., 2018) was used to scaffold the remaining contigs using the 10X data. Scaffolds were polished using Arrow (<https://github.com/PacificBiosciences/gcpp>) and Freebayes (Garrison & Marth, 2012), followed by a manual filtering step to obtain the final nuclear assembly.

Repeats in the assembly were masked using RepeatMasker (Smit et al., 2015) in combination with RepeatModeler2 (Flynn et al., 2020) before starting the annotation procedure. Nuclear genes were annotated using EvidenceModeler (Haas et al., 2008) to generate consensus models of *ab initio* gene predictions, alignments of proteins from closely and distantly related plant species, and transcripts assembled from RNA-seq data. These models were manually filtered to obtain a final set of protein-coding genes.

Used datasets for comparative genome analyses

We mainly focused the comparative genome analyses on *H. incana*, *B. nigra*, and *B. rapa*, three species of the Brassicaceae tribe, of which all members underwent an ancient genome triplication (Lysak et al., 2005; The Brassica rapa Genome Sequencing Project Consortium, 2011). For comparative gene analyses, we extended this group with the Brassicaceae species *A. thaliana*, *A. arabicum*, *S. irio*, *R. raphanistrum*, and *R. sativus*. The latter two *Raphanus* species are also part of the Brassicaceae tribe. Version numbers and locations of all genomes are listed in Table S14.

Analysis of pairwise gene synteny and long terminal repeat retrotransposons in *H. incana*, *B. rapa*, and *B. nigra*

Analyses of pairwise gene synteny between scaffolds of *H. incana* and chromosomes of *B. rapa* and *B. nigra* were performed using the JCVI library (<https://github.com/tanghaibao/jcvi>) (v1.0.5) in Python. Orthologues were identified through all-versus-all alignment of genes with LAST (Kielbasa et al., 2011), retaining reciprocal best hits only (C-score of at least 0.99). Hits were filtered for tandem duplicates (hits located within 10 genes from each other) and chained using the Python implementation of MCScan (Tang et al., 2008) to obtain collinear blocks containing at least four pairs of syntenic genes. Visualisations of collinearity between genomic assemblies were generated using custom scripts of JCVI.

K_s values of syntenic gene pairs were computed using the k_s module of JCVI. Protein sequences of pairs were aligned against each other using MUSCLE (v3.8.1) (Edgar, 2004), after which PAL2NAL (v14) (Suyama et al., 2006) was used to convert protein alignments to nucleotide ones. K_s values for each pair were computed from the nucleotide alignments using the method of Yang & Nielsen, 2000 implemented in PAML (Yang, 2007) (v4.9). Times of divergence between species were estimated by dividing the median of the distributions of their K_s values by the rate of 8.22×10^{-9} synonymous substitutions per year that was established for Brassicaceae species based on extrapolation from the ancient triplication event in the Brassica clade (Beilstein et al., 2010).

Putative LTR-RTs were identified using LTRharvest (v1.6.1) (Ellinghaus et al., 2008) and LTR_finder (v1.1) (Z. Xu & Wang, 2007), after which LTR_retriever (v2.9.0) (Ou & Jiang, 2018) was run with default parameters to filter and combine the output of both tools into a high-confidence set. LTR_retriever was also used to provide estimates of the insertion time of each LTR-RT. Parameters of LTRharvest and LTR_finder were set as recommended in the LTR_retriever documentation. Centromeric regions of the *B. nigra* assembly were obtained from Table S21 of the manuscript describing the assembly (Perumal et al., 2020).

Phylogenetic analysis of *H. incana*, *B. rapa*, and *B. nigra*

The longest isoforms of the nuclear genes of *H. incana*, *B. rapa*, *B. nigra*, and *A. thaliana* (outgroup) were provided to Orthofinder (version 2.3.11) (Emms & Kelly, 2019) to generate phylogenetic species trees. Orthofinder was run using the multiple sequence alignment workflow with default parameters. The same analysis was performed using chloroplast genes. Trees were visualised using iTOL (version 6.3) (Letunic & Bork, 2021).

Comparative Gene Ontology analysis of eight Brassicaceae species

The longest isoforms of the genes of all eight Brassicaceae species described in the section 'Used datasets for comparative genome analyses' were extracted using AGAT (version 0.2.3) (<https://github.com/NBISweden/AGAT>) and clustered into homology groups using the 'group' function of Pantools version 2 (Sheikhzadeh Anari et al., 2018) with a relaxation parameter of 4. Groups were assigned GO slim terms of their associated *A. thaliana* genes (obtained from https://www.arabidopsis.org/download_files/GO_and_PO_Annotations/Gene_Ontology_Annotations/TAIR_GO_slim_categories.txt; last updated on 1 July, 2020) and GO terms assigned to protein domains of associated *H. incana*, *B. rapa*, and *B. nigra* genes using InterProScan (version 5.45–72.0) (Jones et al., 2014) (ran using the Pfam and Panther databases only). GO term enrichment tests were performed using the Fisher exact test and the Benjamini–Hochberg method for multiple testing correction (Benjamini & Hochberg, 1995). *Arabidopsis thaliana* genes were considered to be involved in photosynthesis if they fulfilled one of the following conditions:

- annotated with one of the following GO terms: 'photosynthesis', 'electron transporter, transferring electrons within the cyclic electron transport pathway of photosynthesis activity', or 'electron transporter, transferring electrons within the noncyclic electron transport pathway of photosynthesis activity';
- included in the KEGG pathways ath00195 (Photosynthesis), ath00710 (Carbon fixation in photosynthetic organisms), or ath00196 (Photosynthesis - Antenna Proteins);
- protein products have been assigned the keyword 'Photosynthesis' in the Swiss-Prot database.

The same criteria were used to retain photosynthesis-related genes of *H. incana* while filtering the gene annotation of the assembly (Methods S1).

Investigating the mode of duplicated genes in *H. incana*

Dupgen_finder (GitHub commit hash 8001838) (Qiao et al., 2019) was run with default parameters to determine the mode of duplication for duplicated gene pairs in *H. incana*, using the genome of *A. thaliana* as an outgroup to detect pairs duplicated through whole-genome duplication. Pairs were allowed to be assigned to a single category only. Input files containing alignments of the protein sequences of *H. incana* aligned to themselves and those of *A. thaliana* were prepared using DIAMOND (version 0.9.14) (Buchfink et al., 2015).

Analysis of gene expression under high and low irradiance

Seeds of *H. incana* 'Nijmegen', *B. nigra* 'DG2', *B. rapa* 'R-o-18', and *A. thaliana* Col-0 were germinated on a peat-based potting mixture for nine days under an irradiance of $200 \mu\text{mol m}^{-2} \text{sec}^{-1}$. Twelve seedlings per species were then transferred to 2-L pots filled with a peat-based potting mixture enriched with perlite and 2.5 g L^{-1} Osmocote® Exact Standard 5-6M slow-release fertiliser (ICL Specialty Fertilizers, Geldermalsen, the Netherlands).

Plants were germinated and grown in a climate chamber with a 12/12-h photoperiod at day and night temperatures of 23°C and 20°C, respectively. Relative humidity and CO₂ levels were set at 70% and 400 ppm, respectively. The chamber was equipped with high-output LED light modules (VYPR2p, Fluence by OSRAM, Austin, USA). Six plants per species were assigned to high-light (HL) treatment ($1800 \mu\text{mol m}^{-2} \text{sec}^{-1}$) and the remaining six were assigned to low-light (LL) treatment ($200 \mu\text{mol m}^{-2} \text{sec}^{-1}$). Irradiance uniformity was very high for both HL and LL treatments, with a U2 value of 0.93. Plant positions were randomised across growing areas. Plants were watered with the same custom nutrient solution as the one used in the measurements of photosynthesis rates, daily for LL treatment and twice a day for HL treatment.

Twenty-eight days after sowing, one young fully adapted leaf from each plant was selected, excised, and snap-frozen in liquid nitrogen. Leaf samples were crushed with a mortar and pestle cooled with liquid nitrogen and further homogenised with glass beads for 2 min at 30 Hz in an MM300 Mixer Mill (Retsch GmbH, Haan, Germany). Total RNA was extracted with the RNeasy Plant Mini Kit (QIAGEN N.V., Venlo, the Netherlands) according to the manufacturer's instructions and then subjected to a RQ1 DNase treatment (Promega Corporation, Madison, US) at 37°C for 30 min. We validated the total removal of DNA by means of a no-reverse transcriptase PCR reaction on all RNA samples. RNA was assessed for purity (A_{260}/A_{280}) with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, US) and for possible RNA degradation by means of visual inspection of the RNA on a 1% agarose gel. cDNA was then synthesised from 2 μg total RNA (measured by a spectrophotometer) with the SensiFAST™ cDNA Synthesis Kit (Meridian Bioscience, Cincinnati, US) according to the manufacturer's instructions.

To examine the expression of both single-copy and multi-copy photosynthesis- and/or photoprotection-related genes (Table S15), species-specific RT-qPCR primers were designed with the following criteria: the PCR fragment size had to be between 80 and 120 bp, the maximum difference in melting temperature between primers of the same pair had to be 0.5°C, and overall melting temperatures had to be between 58°C and 62°C. Primers were designed to target a region of the gene as similar as possible in all species. Additionally, for multi-copy genes, the primer pair

had to bind to all copies of a particular gene in one species. RT-qPCR reactions were performed with SYBR green on a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories Inc., Hercules, US). The efficiency of each designed primer set was assessed by means of a standard curve, and only primer sets with efficiencies ranging between 90% and 110% were used. All primer sequences can be found in Table S16.

Gene expression was normalised to the reference genes *ACT2*, *PGK*, *UBQ7*, and *APR* (Joseph et al., 2018; Løvdaal & Lillo, 2009) using the ΔCt method (Livak & Schmittgen, 2001). Normalised gene expression values were calculated as $2^{-\Delta\text{Ct}}$. For the statistical analysis, we performed two-way ANOVA on the ΔCt values with the copy number and light treatment as grouping variables for the multi-copy genes and species and light treatment as grouping variables for the single-copy genes. A Kenward–Roger approximation was used for the degrees of freedom, and a post hoc test was subsequently performed with Tukey's range test, with the significance threshold set at $\alpha = 0.05$.

AUTHOR CONTRIBUTIONS

JH and MGMA initiated the research on the genomic basis of high photosynthesis of *H. incana*. MGMA generated the inbred line of *H. incana* 'Nijmegen'. FG grew plants for the gene expression experiment and performed measurements of photosynthesis rates, flow cytometry experiments, and subsequent analyses of experimental data. FRM performed RNA extractions, cDNA synthesis, and qPCR reactions for the gene expression experiment. FG and VC selected genes and designed primers. RB analysed the results of the gene expression experiment. RB and FB performed DNA and RNA extractions for genome sequencing and assembly. RYW and SS generated strategies for the genome assembly and annotation, which were applied by RYW. RYW and IvdH performed comparative analyses of Brassicaceae genomes. RH was involved in genome annotation. MES helped interpreting the results of the comparative analyses and was involved in drafting the manuscript. JH was involved in the design and interpretation of experiments that were performed to measure photosynthesis rates and in drafting the manuscript. FG, RYW, RB, DDR, MGMA, and SS were majorly involved in overall experimental design and preparing the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We gratefully acknowledge Pádraic Flood for collecting the *H. incana* 'Nijmegen' accession and Graham Taylor for propagating it to perform the first gas exchange measurements. We thank Carlos Alonso Blanco for collecting the *H. incana* 'Burgos' accession, and we thank Niccolò Bassetti and Nina Fatouros for collecting the *B. nigra* 'DG2' accession. We thank the Unifarm staff at the Nergena greenhouse and Radix Klima climate chambers unit for taking excellent care of our plants. We thank Steven Driever for providing valuable insights on the modelling of light–response curves, José van de Belt for her help on chromosome counting, and the Applied Bioinformatics group of Wageningen Plant Research for performing RNA sequencing. This work was supported by project grant ALWGR.2015.9 of the Netherlands Organisation of Scientific Research (NWO) to RYW and RB and the EU-

Horizon2020 Research and Innovation projects CropBooster-P (grant number 817690; www.cropbooster-p.eu; JH and MGMA) and CAPITALISE (grant number 862201; JH and MGMA). We are grateful to the NWO and private partners Rijk Zwaan Breeding, Bejo Zaden, Genetwister Technologies, Averis Seeds, C. Meijer, and HZPC Holland for their financial support through this grant. The funding bodies had no role in the design of the study, data collection, data analysis, interpretation of results, and writing of the manuscript.

DATA AVAILABILITY STATEMENT

Raw sequencing data of *H. incana* can be found on the repository of the National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA612790). The genome assembly of *H. incana* has been deposited at DDBJ/ENA/GenBank under accession number JABCM1000000000. The version described in this manuscript is version JABCM1010000000. The genome assembly and annotation files, as well as a Jbrowse instance of the genome, can be accessed at <https://www.bioinformatics.nl/hirschfeldia>.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Key raw data measured with the LI-COR 6400XT.

Table S2. ANOVA on gross photosynthesis rates.

Table S3. Tukey's range test on net and gross photosynthesis rates.

Table S4. Respiration rates in the dark.

Table S5. Flow cytometry-based haploid genome size estimation.

Table S6. Summary statistics of raw sequencing data.

Table S7. Full QUAST report of nuclear assembly of *H. incana*.

Table S8. Comparing chloroplast assemblies of *H. incana*, *A. thaliana*, *B. rapa*, and *B. nigra*.

Table S9. Alignment of *H. incana* read data against the nuclear assembly.

Table S10. Homology groups of Brassicaceae.

Table S11. Homology groups containing photosynthesis-associated genes.

Table S12. Genes in homology groups expanded in *H. incana*, relative to *A. thaliana*, and found at an equal or higher number of copies in *B. rapa* and *B. nigra*.

Table S13. Enriched GO terms in homology groups expanded in *H. incana*, relative to *A. thaliana*, and found at an equal or higher number of copies in *B. rapa* and *B. nigra*.

Table S14. Used genomic assemblies.

Table S15. *Arabidopsis thaliana* loci of genes of which expression was measured.

Table S16. Primer sequences.

Figure S1. Net CO₂ assimilation rates as measured with the LI-COR 6400XT.

Figure S2. Root tip smears of *H. incana*.

Figure S3. Orthologous syntenic blocks between the genomes of *H. incana* and *B. rapa*.

Figure S4. Normalised relative gene expression per species and light treatment for the set of multi-copy photosynthesis-related genes.

Figure S5. Normalised relative gene expression per species and light treatment for the set of single-copy photosynthesis-related genes.

Figure S6. Histogram of mismatch rates of PacBio reads aligned against the chloroplast and mitochondrial assembly of *Hirschfeldia incana*.

Methods S1. Pre-processing of reads. Identifying nuclear, organellar, and contaminant reads. Assembling the chloroplast genome. Assembling the mitochondrial genome. Assembling the nuclear genome. Genome annotation.

OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA612790>.

REFERENCES

- Anderson, J.K. & Warwick, S.I. (1999) Chromosome number evolution in the tribe Brassiceae (Brassicaceae): evidence from isozyme number. *Plant Systematics and Evolution*, **215**(1–4), 255–285.
- Arias, T. & Pires, J.C. (2012) A fully resolved chloroplast phylogeny of the brassica crops and wild relatives (Brassicaceae: Brassiceae): novel clades and potential taxonomic implications. *Taxon*, **61**(5), 980–988.
- Arumuganathan, K. & Earle, E. (1991) Estimation of nuclear DNA content of plants by flow cytometry. *Plant Molecular Biology Reporter*, **9**(3), 229–241.
- Auguay, F., Fahr, M., Moulin, P., Brugel, A., Laplaze, L., El Mzibri, M. et al. (2013) Lead tolerance and accumulation in *Hirschfeldia incana*, a mediterranean brassicaceae from metalliferous mine spoils. *PLoS One*, **8**(5), e61932.
- Auguay, F., Fahr, M., Moulin, P., El Mzibri, M., Smouni, A., Filali-Maltouf, A. et al. (2016) Transcriptome changes in *Hirschfeldia incana* in response to lead exposure. *Frontiers in Plant Science*, **6**, 1231.
- Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Loudet, O. et al. (2021) Genetic and environmental modulation of transposition shapes the evolutionary potential of *Arabidopsis thaliana*. *Genome Biology*, **22**(1), 138.
- Bagheri, H., El-Soda, M., van Oorschot, I., Hanhart, C., Bonnema, G., Jansen-van den Bosch, T. et al. (2012) Genetic analysis of morphological traits in a new, versatile, rapid-cycling *Brassica rapa* recombinant inbred line population. *Frontiers in Plant Science*, **3**, 183.
- Bancroft, I., Morgan, C., Fraser, F., Higgins, J., Wells, R., Clissold, L. et al. (2011) Dissecting the genome of the polyploid crop oilseed rape by transcriptome sequencing. *Nature Biotechnology*, **29**(8), 762–766.
- Beilstein, M.A., Nagalingum, N.S., Clements, M.D., Manchester, S.R. & Mathews, S. (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences*, **107**(43), 18724–18728.
- Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F.-C., Falentin, C. et al. (2018) Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*, **4**(11), 879–887.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**(1), 289–300.
- Bennett, M.D., Leitch, I.J., Price, H.J. & Johnston, J.S. (2003) Comparisons with *Caenorhabditis* (~100 mb) and *Drosophila* (~175 mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of ~125 mb. *Annals of Botany*, **91**(5), 547–557.
- Beric, A., Mabry, M.E., Harkess, A.E., Brose, J., Schranz, M.E., Conant, G.C. et al. (2021) Comparative phylogenetics of repetitive elements in a diverse order of flowering plants (Brassicales). *G3: Genes | Genomes | Genetics*, **11**(7), jkab14.
- Buchfink, B., Xie, C. & Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**(1), 59–60.
- Cai, X., Wu, J., Liang, J., Lin, R., Zhang, K., Cheng, F. et al. (2020) Improved *Brassica oleracea* JZS assembly reveals significant changing of LTR-RT dynamics in different morphotypes. *Theoretical and Applied Genetics*, **133**(11), 3187–3199.
- Canvin, D.T., Berry, J.A., Badger, M.R., Fock, H. & Osmond, C.B. (1980) Oxygen exchange in leaves in the light. *Plant Physiology*, **66**(2), 302–307.
- Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X. et al. (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**(6199), 950–953.
- Chang, S., Puryear, J. & Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter*, **11**(2), 113–116.
- Choi, S.R., Teakle, G.R., Plaha, P., Kim, J.H., Allender, C.J., Beynon, E. et al. (2007) The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project. *Theoretical and Applied Genetics*, **115**(6), 777–792.
- Coombe, L., Zhang, J., Vandervalk, B.P., Chu, J., Jackman, S.D., Birol, I. et al. (2018) ARKS: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinformatics*, **19**(1), 1–10.
- Crafts-Brandner, S.J. & Salvucci, M.E. (2002) Sensitivity of photosynthesis in a C4 plant, maize, to heat stress. *Plant Physiology*, **129**(4), 1773–1780.
- Crocco, C.D., Ocampo, G.G., Ploschuk, E.L., Mantese, A. & Botto, J.F. (2018) Heterologous expression of *AtBBX21* enhances the rate of photosynthesis and alleviates photoinhibition in *Solanum tuberosum*. *Plant Physiology*, **177**(1), 369–380.
- Darmency, H. & Fleury, A. (2000) Mating system in *Hirschfeldia incana* and hybridization to oilseed rape. *Weed Research (Oxford)*, **40**(2), 231–238.
- Dassanayake, M., Oh, D.-H., Haas, J.S., Hernandez, A., Hong, H., Ali, S. et al. (2011) The genome of the extremophile crucifer *Thellungiella parvula*. *Nature Genetics*, **43**(9), 913–918.
- De Souza, A.P., Burgess, S.J., Doran, L., Hansen, J., Manukyan, L., Maryn, N. et al. (2022) Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. *Science*, **377**(6608), 851–854.
- Driever, S., Lawson, T., Andralojc, P., Raines, C. & Parry, M. (2014) Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. *Journal of Experimental Botany*, **65**(17), 4959–4973.
- Du, J., Tian, Z., Sui, Y., Zhao, M., Song, Q., Cannon, S.B. et al. (2012) Pericentromeric effects shape the patterns of divergence, retention, and expression of duplicated genes in the paleopolyploid soybean. *The Plant Cell*, **24**(1), 21–32.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5), 1792–1797.
- Edger, P.P. & Pires, J.C. (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Research*, **17**(5), 699–717.
- Edwards, E.J., Osborne, C.P., Strömberg, C.A., Smith, S.A. & C4 Grasses Consortium. (2010) The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science*, **328**(5978), 587–591.
- Ehleringer, J. (1985) Annuals and perennials of warm deserts. In: Chabot, B.H. & Mooney, H.A. (Eds.) *Physiological Ecology of North American Plant Communities*. Dordrecht: Springer, pp. 162–180.
- Ellinghaus, D., Kurtz, S. & Willhoeft, U. (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**(1), 18.
- Emms, D.M. & Kelly, S. (2019) Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**(1), 1–14.
- Fahr, M., Laplaze, L., El Mzibri, M., Doumas, P., Bendaou, N., Hoher, V. et al. (2015) Assessment of lead tolerance and accumulation in metallicolour and non-metallicolour populations of *Hirschfeldia incana*. *Environmental and Experimental Botany*, **109**, 186–192.
- Favory, J.-J., Kobayashi, M., Tanaka, K., Peltier, G., Kreis, M., Valay, J.-G. et al. (2005) Specific function of a plastid sigma factor for *ndh f* gene transcription. *Nucleic Acids Research*, **33**(18), 5991–5999.
- Flood, P.J., Harbinson, J. & Aarts, M.G.M. (2011) Natural genetic variation in plant photosynthesis. *Trends in Plant Science*, **16**(6), 327–335.
- Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C. et al. (2020) Repeatmodeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, **117**(17), 9451–9457.

- Franzke, A., German, D., Al-Shehbaz, I.A. & Mummenhoff, K. (2009) Arabidopsis family ties: molecular phylogeny and age estimates in Brassicaceae. *Taxon*, **58**(2), 425–437.
- Furbank, R.T., Jimenez-Berni, J.A., George-Jaeggli, B., Potgieter, A.B. & Deery, D.M. (2019) Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytologist*, **223**(4), 1714–1727.
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*. Preprint at <https://arxiv.org/abs/1207.3907>.
- Genty, B. & Harbinson, J. (1996) Regulation of light utilization for photosynthetic electron transport. In: Baker, N.R. (Ed.) *Photosynthesis and the environment*. Dordrecht: Springer, pp. 67–99.
- Gibson, A.C. (1998) Photosynthetic organs of desert plants. *Bioscience*, **48** (11), 911–920.
- Gilbert, M.E., Zwieniecki, M.A. & Holbrook, N.M. (2011) Independent variation in photosynthetic capacity and stomatal conductance leads to differences in intrinsic water use efficiency in 11 soybean genotypes before and during mild drought. *Journal of Experimental Botany*, **62**(8), 2875–2887.
- Gitelson, A.A., Peng, Y., Arkebauer, T.J. & Suyker, A.E. (2015) Productivity, absorbed photosynthetically active radiation, and light use efficiency in crops: implications for remote sensing of crop primary production. *Journal of Plant Physiology*, **177**, 100–109.
- Gu, J., Yin, X., Stomph, T.-J. & Struik, P.C. (2014) Can exploiting natural genetic variation in leaf photosynthesis contribute to increasing rice productivity? A simulation analysis. *Plant, Cell & Environment*, **37**(1), 22–34.
- Gu, J., Yin, X., Struik, P.C., Stomph, T.J. & Wang, H. (2012) Using chromosome introgression lines to map quantitative trait loci for photosynthesis parameters in rice (*Oryza sativa* L.) leaves under drought and well-watered field conditions. *Journal of Experimental Botany*, **63**(1), 455–469.
- Gu, J., Zhou, Z., Li, Z., Chen, Y., Wang, Z., Zhang, H. *et al.* (2017) Photosynthetic properties and potentials for improvement of photosynthesis in pale green leaf rice under high light conditions. *Frontiers in Plant Science*, **8**, 1082.
- Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y. & Durbin, R. (2020) Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, **36**(9), 2896–2898.
- Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**(8), 1072–1075.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J. *et al.* (2008) Automated eukaryotic gene structure annotation using Evidence-Modeler and the program to assemble spliced alignments. *Genome Biology*, **9**(1), R7.
- Harbinson, J., Kaiser, E. & Morales, A.S. (2022) Integrating the stages of photosynthesis. In: Ruban, A., Foyer, C.H. & Murchie, E.H. (Eds.) *Photosynthesis in action*. London: Elsevier, pp. 195–242.
- He, Z., Ji, R., Havlickova, L., Wang, L., Li, Y., Lee, H.T. *et al.* (2021) Genome structural evolution in brassica crops. *Nature Plants*, **7**(6), 757–765.
- Heddad, M., Norén, H., Reiser, V., Dunaeva, M., Andersson, B. & Adamska, I. (2006) Differential expression and localization of early light-induced proteins in Arabidopsis. *Plant Physiology*, **142**(1), 75–87.
- Huang, C.-H., Sun, R., Hu, Y., Zeng, L., Zhang, N., Cai, L. *et al.* (2016) Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. *Molecular Biology and Evolution*, **33**(2), 394–412.
- Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K. & Havaux, M. (2003) Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proceedings of the National Academy of Sciences*, **100**(8), 4921–4926.
- Johnston, J.S., Pepper, A.E., Hall, A.E., Chen, Z.J., Hodnett, G., Drabek, J. *et al.* (2005) Evolution of genome size in Brassicaceae. *Annals of Botany*, **95**(1), 229–235.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C. *et al.* (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**(9), 1236–1240.
- Joseph, J.T., Poolakkalody, N.J. & Shah, J.M. (2018) Plant reference genes for development and stress response studies. *Journal of Biosciences*, **43** (1), 173–187.
- Katche, E., Quezada-Martinez, D., Katche, E.I., Vasquez-Teuber, P. & Mason, A.S. (2019) Interspecific hybridization for brassica crop improvement. *Crop Breeding, Genetics and Genomics*, **1**(1), e190007.
- Kent, T.V., Uzunović, J. & Wright, S.I. (2017) Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **372**(1736), 20160458.
- Kielbasa, S.M., Wan, R., Sato, K., Horton, P. & Frith, M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Research*, **21**(3), 487–493.
- Kim, H., Choi, S.R., Bae, J., Hong, C.P., Lee, S.Y., Hossain, M.J. *et al.* (2009) Sequenced BAC anchored reference genetic map that reconciles the ten individual chromosomes of *Brassica rapa*. *BMC Genomics*, **10**(1), 432.
- Koornneef, M. & Meinke, D. (2010) The development of Arabidopsis as a model plant. *The Plant Journal*, **61**(6), 909–921.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H. & Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, **27**(5), 722–736.
- Lawson, T., Kramer, D.M. & Raines, C.A. (2012) Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Current Opinion in Biotechnology*, **23**(2), 215–220.
- Leakey, A.D., Uribealarea, M., Ainsworth, E.A., Naidu, S.L., Rogers, A., Ort, D.R. *et al.* (2006) Photosynthesis, productivity, and yield of maize are not affected by open-air elevation of CO₂ concentration in the absence of drought. *Plant Physiology*, **140**(2), 779–790.
- Lee, P.L., Patel, R., Conlan, R., Wainwright, S. & Hipkin, C. (2004) Comparison of genetic diversities in native and alien populations of hoary mustard (*Hirschfeldia incana* [L.] Lagreze-Fossat). *International Journal of Plant Sciences*, **165**(5), 833–843.
- Lee, S.-K., Jeon, J.-S., Boernke, F., Voll, L., Cho, J.-I., Goh, C.-H. *et al.* (2008) Loss of cytosolic fructose-1, 6-bisphosphatase limits photosynthetic sucrose synthesis and causes severe growth retardations in rice (*Oryza sativa*). *Plant, Cell & Environment*, **31**(12), 1851–1863.
- Leister, D. (2019) Thawing out frozen metabolic accidents. *BMC Biology*, **17** (1), 1–11.
- Letunic, I. & Bork, P. (2021) Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, **49** (W1), W293–W296.
- Li, Y., Liu, B., Zhang, J., Kong, F., Zhang, L., Meng, H. *et al.* (2019) OHP1, OHP2, and HCF244 form a transient functional complex with the photosystem II reaction center. *Plant Physiology*, **179**(1), 195–208.
- Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y. & De Smet, R. (2016) Gene duplicability of core genes is highly consistent across all angiosperms. *The Plant Cell*, **28**(2), 326–344.
- Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A. *et al.* (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nature Communications*, **5**(1), 1–11.
- Liu, Y., Wei, W., Ma, K., Li, J., Liang, Y. & Darmency, H. (2013) Consequences of gene flow between oilseed rape (*Brassica napus*) and its relatives. *Plant Science*, **211**, 42–51.
- Livak, K.J. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2^{-ΔΔCt} method. *Methods*, **25**(4), 402–408.
- Long, S.P., Zhu, X.-G., Naidu, S.L. & Ort, D.R. (2006) Can improvement in photosynthesis increase crop yields? *Plant, Cell & Environment*, **29**(3), 315–330.
- Løvdaal, T. & Lillo, C. (2009) Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Analytical Biochemistry*, **387**(2), 238–242.
- Lysak, M.A., Koch, M.A., Pecinka, A. & Schubert, I. (2005) Chromosome triplication found across the tribe Brassicaceae. *Genome Research*, **15**(4), 516–525.
- Maiwald, D., Dietzmann, A., Jahns, P., Pesaresi, P., Joliot, P., Joliot, A. *et al.* (2003) Knock-out of the genes coding for the Rieske protein and the ATP-synthase δ -subunit of Arabidopsis. Effects on photosynthesis, thylakoid protein composition, and nuclear chloroplast gene expression. *Plant Physiology*, **133**(1), 191–202.
- Mandáková, T., Li, Z., Barker, M.S. & Lysak, M.A. (2017) Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *The Plant Journal*, **91**(1), 3–21.

- Mandáková, T. & Lysak, M.A. (2018) Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology*, **42**, 55–65.
- Marçais, G. & Kingsford, C. (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**(6), 764–770.
- Marshall, B. & Biscoe, P. (1980) A model for C3 leaves describing the dependence of net photosynthesis on irradiance. *Journal of Experimental Botany*, **31**(1), 29–39.
- Martinis, J., Glauser, G., Valimareanu, S. & Kessler, F. (2013) A chloroplast ABC1-like kinase regulates vitamin E metabolism in *Arabidopsis*. *Plant Physiology*, **162**(2), 652–662.
- Mira, S., Veiga-Barbosa, L. & Pérez-García, F. (2019) Seed dormancy and longevity variability of *Hirschfeldia incana* L. during storage. *Seed Science Research*, **29**(2), 97–103.
- Monneveux, P., Pastenes, C. & Reynolds, M.P. (2003) Limitations to photosynthesis under light and heat stress in three high-yielding wheat genotypes. *Journal of Plant Physiology*, **160**(6), 657–666.
- Munekage, Y., Hashimoto, M., Miyake, C., Tomizawa, K.-I., Endo, T., Tasaka, M. *et al.* (2004) Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*, **429**(6991), 579–582.
- Murchie, E.H., Chen, Y.-Z., Hubbart, S., Peng, S. & Horton, P. (1999) Interactions between senescence and leaf orientation determine in situ patterns of photosynthesis and photoinhibition in field-grown rice. *Plant Physiology*, **119**(2), 553–564.
- Norén, H., Svensson, P., Stegmark, R., Funk, C., Adamska, I. & Andersson, B. (2003) Expression of the early light-induced protein but not the PsbS protein is influenced by low temperature and depends on the developmental stage of the plant in field-grown pea cultivars. *Plant, Cell & Environment*, **26**(2), 245–253.
- Oh, D.-H., Dassanayake, M., Bohnert, H.J. & Cheeseman, J.M. (2013) Life at the extreme: lessons from the genome. *Genome Biology*, **13**(3), 1–9.
- Ortiz, D., Hu, J. & Salas Fernandez, M.G. (2017) Genetic architecture of photosynthesis in *Sorghum bicolor* under non-stress and cold stress conditions. *Journal of Experimental Botany*, **68**(16), 4545–4557.
- Ou, S. & Jiang, N. (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**(2), 1410–1422.
- Paritosh, K., Pradhan, A.K. & Pental, D. (2020) A highly contiguous genome assembly of *Brassica nigra* (BB) and revised nomenclature for the pseudochromosomes. *BMC Genomics*, **21**(1), 1–12.
- Peng, L., Fukao, Y., Fujiwara, M., Takami, T. & Shikanai, T. (2009) Efficient operation of NAD(P)H dehydrogenase requires supercomplex formation with photosystem I via minor LHCl in *Arabidopsis*. *The Plant Cell*, **21**(11), 3623–3640.
- Perumal, S., Koh, C.S., Jin, L., Buchwaldt, M., Higgins, E.E., Zheng, C. *et al.* (2020) A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *brassica* genome. *Nature Plants*, **6**(8), 929–941.
- Pleban, J.R., Mackay, D.S., Aston, T.L., Ewers, B.E. & Weinig, C. (2018) Phenotypic trait identification using a multimodel Bayesian method: a case study using photosynthesis in *Brassica rapa* genotypes. *Frontiers in Plant Science*, **9**, 448.
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R. *et al.* (2019) Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biology*, **20**(1), 1–23.
- Ritz, K.R., Noor, M.A. & Singh, N.D. (2017) Variation in recombination rate: adaptive or not? *Trends in Genetics*, **33**(5), 364–374.
- Rizzon, C., Ponger, L. & Gaut, B.S. (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLoS Computational Biology*, **2**(9), e115.
- Sánchez-Yélamo, M. (2009) Relationships in the *Diplotaxis–Erucastrum–Brassica* complex (Brassicaceae) evaluated from isoenzymatic profiles of the accessions as a whole. Applications for characterisation of phylogenetic resources preserved *ex situ*. *Genetic Resources and Crop Evolution*, **56**(7), 1023–1036.
- Schranz, M.E., Lysak, M.A. & Mitchell-Olds, T. (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends in Plant Science*, **11**(11), 535–542.
- Sheikhzadeh Anari, S., de Ridder, D., Schranz, M.E. & Smit, S. (2018) Efficient inference of homologs in large eukaryotic pan-proteomes. *BMC Bioinformatics*, **19**(1), 340.
- Shi, T., Bibby, T.S., Jiang, L., Irwin, A.J. & Falkowski, P.G. (2005) Protein interactions limit the rate of evolution of photosynthetic genes in cyanobacteria. *Molecular Biology and Evolution*, **22**(11), 2179–2189.
- Siemens, J. (2011) Hirschfeldia. In: Kole, C. (Ed.) *Wild Crop Relatives: Genomic and Breeding Resources*. Heidelberg: Springer, pp. 171–176.
- Smit, A. F., Hubley, R., & Green, P. (2015). *Repeatmasker open-4.0*. <http://www.repeatmasker.org> 2013-2015
- Somerville, C. & Koornneef, M. (2002) A fortunate choice: the history of *Arabidopsis* as a model plant. *Nature Reviews Genetics*, **3**(11), 883–889.
- Stephenson, P., Baker, D., Girin, T., Perez, A., Amoah, S., King, G.J. *et al.* (2010) A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC Plant Biology*, **10**(1), 1–10.
- Stinziano, J. R., Roback, C., Gamble, D., Murphy, B., Hudson, P., & Muir, C. D. (2020). Photosynthesis: Tools for plant ecophysiology & modeling [R package version 2.0.1]. <https://CRAN.R-project.org/package=photosynthesis>
- Strigens, A., Freitag, N.M., Gilbert, X., Grieder, C., Riedelsheimer, C., Schrag, T.A. *et al.* (2013) Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. *Plant, Cell & Environment*, **36**(10), 1871–1887.
- Sun, H., Ding, J., Piednoël, M. & Schneeberger, K. (2018) findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics*, **34**(4), 550–557.
- Suryawanshi, V., Talke, I.N., Weber, M., Eils, R., Brors, B., Clemens, S. *et al.* (2016) Between-species differences in gene copy number are enriched among functions critical for adaptive evolution in *Arabidopsis halleri*. *BMC Genomics*, **17**(13), 43–64.
- Suyama, M., Torrents, D. & Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**(suppl_2), W609–W612.
- Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M. & Paterson, A.H. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research*, **18**(12), 1944–1954.
- Taylor, S.H., Orr, D.J., Carmo-Silva, E. & Long, S.P. (2020) During photosynthetic induction, biochemical and stomatal limitations differ between Brassica crops. *Plant, Cell & Environment*, **43**(11), 2623–2636.
- The Brassica rapa Genome Sequencing Project Consortium. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics*, **43**(10), 1035–1039.
- Tian, Z., Rizzon, C., Du, J., Zhu, L., Bennetzen, J.L., Jackson, S.A. *et al.* (2009) Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? *Genome Research*, **19**(12), 2221–2230.
- Tsunoyama, Y., Ishizaki, Y., Morikawa, K., Kobori, M., Nakahira, Y., Takeba, G. *et al.* (2004) Blue light-induced transcription of plastid-encoded *psbD* gene is mediated by a nuclear-encoded transcription initiation factor, atsig5. *Proceedings of the National Academy of Sciences*, **101**(9), 3304–3309.
- Turner, D.P., Urbanski, S., Bremer, D., Wofsy, S.C., Meyers, T., Gower, S.T. *et al.* (2003) A cross-biome comparison of light use efficiency for gross primary production. *Global Change Biology*, **9**(3), 383–395.
- van Bezouw, R.F., Keurentjes, J.J., Harbinson, J. & Aarts, M.G.M. (2019) Converging phenomics and genomics to study natural variation in plant photosynthetic efficiency. *The Plant Journal*, **97**(1), 112–133.
- van Rooijen, R., Aarts, M.G.M. & Harbinson, J. (2015) Natural genetic variation for acclimation of photosynthetic light use efficiency to growth irradiance in *Arabidopsis*. *Plant Physiology*, **167**(4), 1412–1429.
- van Rooijen, R., Kruijjer, W., Boesten, R., Van Eeuwijk, F.A., Harbinson, J. & Aarts, M.G.M. (2017) Natural variation of yellow seedling1 affects photosynthetic acclimation of *Arabidopsis thaliana*. *Nature Communications*, **8**(1), 1–9.
- von Caemmerer, S. & Evans, J.R. (2010) Enhancing C3 photosynthesis. *Plant Physiology*, **154**(2), 589–592.
- Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J. *et al.* (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**(14), 2202–2204.
- Wang, X., Torres, M.J., Pierce, G., Lemke, C., Nelson, L.K., Yuksel, B. *et al.* (2011) A physical map of *Brassica oleracea* shows complexity of chromosomal changes following recursive paleopolyploidizations. *BMC Genomics*, **12**(1), 1–15.

- Werk, K., Ehleringer, J., Forseth, I. & Cook, C. (1983) Photosynthetic characteristics of Sonoran Desert winter annuals. *Oecologia*, **59**(1), 101–105.
- Xu, C., Yinan, C., Zhang, L., Jian, W., Liang, J., Cheng, L. *et al.* (2018) Hot-spots of independent and multiple rounds of LTR-retrotransposon bursts in brassica species. *Horticultural Plant Journal*, **4**(4), 165–174.
- Xu, Z. & Wang, H. (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research*, **35**(suppl_2), W265–W268.
- Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8), 1586–1591.
- Yang, Z. & Nielsen, R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, **17**(1), 32–43.
- Youssef, A., Laizet, Y., Block, M.A., Maréchal, E., Alcaraz, J.-P., Larson, T.R. *et al.* (2010) Plant lipid-associated fibrillin proteins condition jasmonate production under photosynthetic stress. *The Plant Journal*, **61**(3), 436–445.
- Zhang, L., Cai, X., Wu, J., Liu, M., Grob, S., Cheng, F. *et al.* (2018) Improved *Brassica rapa* reference genome by single-molecule sequencing and chromosome conformation capture technologies. *Horticulture Research*, **5**(1), 1–11.
- Zhao, M., Du, J., Lin, F., Tong, C., Yu, J., Huang, S. *et al.* (2013) Shifts in the evolutionary rate and intensity of purifying selection between two brassica genomes revealed by analyses of orthologous transposons and relics of a whole genome triplication. *The Plant Journal*, **76**(2), 211–222.
- Zhu, X.-G., Long, S.P. & Ort, D.R. (2010) Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology*, **61**, 235–261.
- Żmienko, A., Samelak, A., Kozłowski, P. & Figlerowicz, M. (2014) Copy number polymorphism in plant genomes. *Theoretical and Applied Genetics*, **127**(1), 1–18.