Plant Science 344 (2024) 112109

Contents lists available at ScienceDirect

Plant Science

journal homepage: www.elsevier.com/locate/plantsci

Different reference genomes determine different results: Comparing SNP calling in RAD-seq of *Engelhardia roxburghiana* using different reference genomes

Pei-Han Huang [a,b,e], Tian-Rui Wang [b,c,e], Min Li [a,e], Ou-Yan Fang [a,e], Ren-Ping Su [a,e], Hong-Hu Meng [a,d,*], Yi-Gang Song [b,**], Jie Li [a,**]

[a] *Plant Phylogenetics and Conservation Group, Center for Integrative Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China*
[b] *Eastern China Conservation Centre for Wild Endangered Plant Resources, Shanghai Chenshan Botanical Garden, Shanghai, 201602, China*
[c] *Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China*
[d] *Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Nay Pyi Taw 05282, Myanmar*
[e] *University of Chinese Academy of Sciences, Beijing 100049, China*

ARTICLE INFO

Keywords:
RAD-seq
STACKS
SNP calling
Reference-based approach

ABSTRACT

Advances in next-generation sequencing (NGS) have significantly reduced the cost and improved the efficiency of obtaining single nucleotide polymorphism (SNP) markers, particularly through restriction site-associated DNA sequencing (RAD-seq). Meanwhile, the progression in whole genome sequencing has led to the utilization of an increasing number of reference genomes in SNP calling processes. This study utilized RAD-seq data from 242 individuals of *Engelhardia roxburghiana*, a tropical tree of the walnut family (Juglandaceae), with SNP calling conducted using the STACKS pipeline. We aimed to compare both reference-based approaches, namely, employing a closely related species as the reference genome versus the species itself as the reference genome, to evaluate their respective merits and limitations. Our findings indicate a substantial discrepancy in the number of obtained SNPs between using a closely related species as opposed to the species itself as reference genomes, the former yielded approximately an order of magnitude fewer SNPs compared to the latter. While the missing rate of individuals and sites of the final SNPs obtained in the two scenarios showed no significant difference. The results showed that using the reference genome of the species itself tends to be prioritized in RAD-seq studies. However, if this is unavailable, considering closely related genomes is feasible due to their wide applicability and low missing rate as alternatives. This study contributes to enrich the understanding of the impact of SNP acquisition when utilizing different reference genomes.

The emergence of next-generation sequencing (NGS) techniques has profoundly influenced life science researches (McGinn and Gut, 2013; Gibbs, 2020; Uhlen and Quake, 2023), which can be sequenced with high-throughput, strong-scalability, low-cost and fast-speed (Hudson, 2008). Against this background, various genome sequencing approaches have been invented to identify and genotype thousands of markers for genomic screening. Among these approaches, restriction site-associated DNA sequencing (RAD-seq) has emerged as a widely adopted method for single nucleotide polymorphism (SNP) discovery and genotyping, especially in the studies of non-model organisms (Davey et al., 2011; Andrews et al., 2016). A typical RAD-seq project involves several stages, including sample collection, DNA extraction, RAD library construction using restriction enzymes (REs) to determine the set of loci to be sequenced, analysis of the resulting short fragments, and eventually conducting analyses based on the acquired information (Baird et al., 2008; Etter et al., 2011). Over the past decade, the proliferation of RAD-seq applications significantly propelled researches in ecological, evolutionary, and conservation genomics (Wagner et al., 2013; Pante et al., 2015; Orita et al., 2021; Hsu et al., 2022; Probowati et al., 2023), owing to the abundance of genetic markers identified and genotyped in a

* Corresponding author at: Plant Phylogenetics and Conservation Group, Center for Integrative Conservation & Yunnan Key Laboratory for Conservation of Tropical Rainforests and Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China.
** Corresponding authors.
E-mail addresses: menghonghu@xtbg.ac.cn (H.-H. Meng), ygsong@cemps.ac.cn (Y.-G. Song), jieli@xtbg.ac.cn (J. Li).

https://doi.org/10.1016/j.plantsci.2024.112109
Received 8 January 2024; Received in revised form 23 April 2024; Accepted 30 April 2024
Available online 3 May 2024

single step. Additionally, many different RAD variations, such as ddRAD-seq (Peterson et al., 2012), 2bRAD-seq (Wang et al., 2012), GBS (Elshire et al., 2011), were developed based on the original RAD-seq methodology, aiming to offer improved strategies for specific scenarios, and these variations have been extensively applied in genetic diversity, phylogeny and biodiversity conservation (e.g., Mu et al., 2020; Duan et al., 2022; Probowati et al., 2023; Su et al., 2023; Zhao et al., 2023)

Meanwhile, multiple bioinformatic pipelines have been developed to process RAD-seq data and identify a vast number of SNPs. Among this, STACKS (Catchen et al., 2011, 2013) and ipyrad (Eaton and Overcast, 2020) are extensively used bioinformatic pipelines for they can address both reference-based approach and de novo approach (i.e., without a reference genome). While STACKS is often employed for population genetics purposes, ipyrad is more commonly used for phylogenetic studies. This study mainly focuses on STACKS, which established protocols across several applications (Rochette and Catchen, 2017). This pipeline designed to assemble loci from short-read sequences derived from restriction enzyme-based protocols (Catchen et al., 2011, 2013). To date, as the advancements in whole genome sequencing assembly technology, an increasing number of species have successfully undergone whole genome assembly. As of January 2024, over 36,000 eukaryotes had been sequenced and cataloged in the NCBI database (NCBI 2024). Although this constitutes only a fraction of the known eukaryotic species, it suggests a tenfold increase from the number reported in 2016 (NCBI 2016). When a species lacks its own reference genome, alternative options like employing reference genomes from its related species (e.g., Paris et al., 2017; Shen et al., 2023) or utilizing a de novo approach (e.g., Su et al., 2023; Piwczyński et al., 2023) have become available for analysis. Both reference genome-based approach and de novo approach possess distinct advantages. Reference genomes play a pivotal role in distinguishing orthologous sites from paralogous sites and correct low-level sequencing errors in reads (Davey and Blaxter, 2010; Rubin et al., 2012). Simultaneously, reference genome-based approaches enable analyses with enhanced statistical power, such as sliding window analysis (e.g., Martin et al., 2013; Ruegg et al., 2014), making them more efficient in terms of time and computational resources compared to the de novo approach. Nevertheless, de novo methods have become prevalent in pan-genome studies (e.g., Wang et al., 2023; Kang et al., 2023). When a single reference genome of a species cannot encompass all its genetic information, a de novo approach can be employed to assemble a pan-genome containing greater genetic diversity, enabling a more profound exploration of the species' deep phylogenetic relationships.

Amidst the rapid advancement of RAD-seq and its associated bioinformatic pipelines, a myriad of evaluations concerning the RAD-seq method have emerged. These encompass performance analyses among various bioinformatic pipelines (Casanova et al., 2021) and comparisons between the de novo approach and reference-based approach (Torkamaneh et al., 2016; Shafer et al., 2017; Dittberner et al., 2018; Casanova et al., 2021). Nonetheless, limited attention has been devoted in prior research to scrutinizing the impact of comparing the use of closely related species as the reference genome versus the species itself as the reference genome. Therefore, evaluating the advantages and limitations of these two strategies will offer some valuable insights for future researchers.

To disentangle this pressing issue, we conducted SNP calling and selected several crucial parameters as filtering standards for this analysis. The RAD-seq data generated from 242 individuals of *Engelhardia roxburghiana* Wall. (= *Alfaropsis* Iljinsk.) (Stone, 2010) in 50 populations (Table S1), an evergreen tree with even-pinnate leaves, orange-red sprouts, dark-brown or black twigs, leaflets typically arranged in 3–5 pairs, most of them having a short acuminate apex and the secondary leaflet veins are in 7 (5−13) pairs (Meng et al., 2022a, b; Zhang et al., 2020). Our discussion revolved around two strategies: employing a closely related species (i.e., *Pterocarya stenoptera*) as the reference genome, and utilizing the species itself (i.e., *E. roxburghiana*) as the reference genome (Both available at: https://cmb.bnu.edu.cn/juglans/). *E. roxburghiana* and *P. stenoptera* belong to different genera of Juglandaceae, and the two species are closely related species (Ding et al., 2023). In terms of morphology, within the Juglandaceae family, fruits bearing fruit wings include *Engelhardia*, *Pterocarya*, and *Cyclocarya*. Additionally, *Cyclocarya* features one fruit wing surrounding the fruit, *Pterocarya* exhibits two fruit wings, while *Engelhardia* possesses three fruit wings (Lu et al., 1999). Therefore, we used the two species as reference genomes to explore the SNP calling in RAD-seq of *E. roxburghiana*.

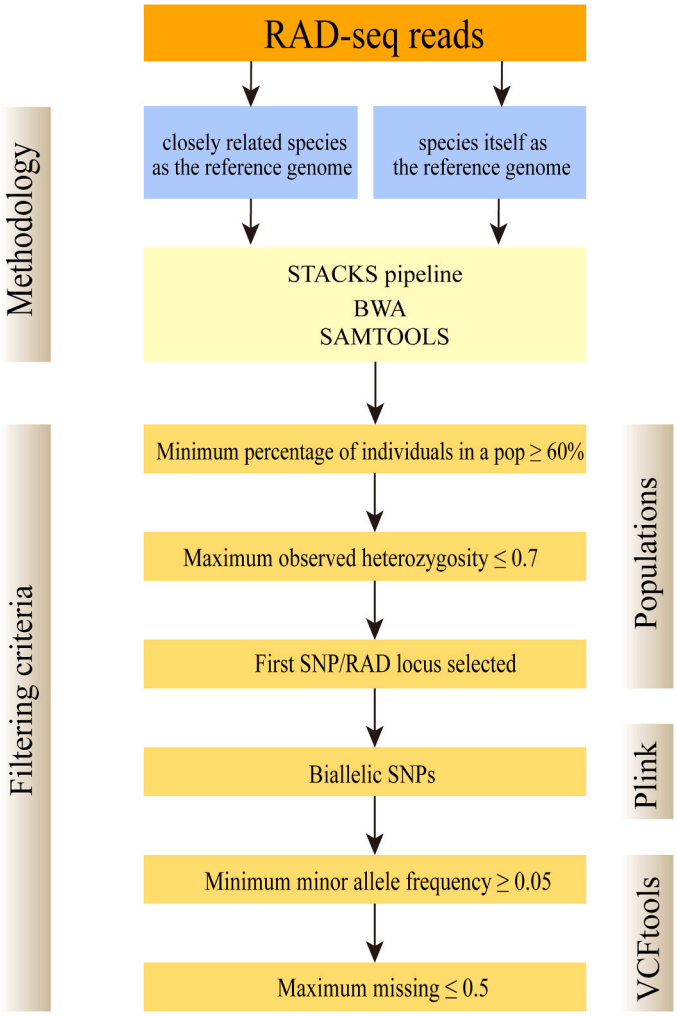

Fig. 1. Overview of this study. The RAD-seq data obtained from a total of 242 individuals of *E. roxburghiana* were analyzed using reference-based methods, including the methodology and filtering criteria (Minimum percentage of individuals in a pop ≥ 60%: -r 0.6; Maximum observed heterozygosity ≤ 0.7: –max-obs-het 0.7; First SNP/RAD locus selected: –write-single-snp; Biallelic SNPs: –biallelic-only; Minimum minor allele frequency ≥ 0.05: –maf 0.05; Maximum missing ≤ 0.5: –max-missing 0.5).

The number of high-quality SNPs remained was used as the evaluation criteria. Initially, we extracted the total genomic DNA of all samples and send them to BGI (Shenzhen, Guangdong, China) for library construction and sequencing. The resulting data remained with high-quality, ensuring its suitability for subsequent analyses (Table S2). STACKS was chosen to perform SNP calling, and reads were filtered for overall quality, demultiplexed, and trimmed to 120 bp. The reference genomes of *P. stenoptera* and *E. roxburghiana* were downloaded, and the average percentage of the RAD reads mapped to these reference genomes achieved 38.93% and 97.09%, respectively (Table S3). We

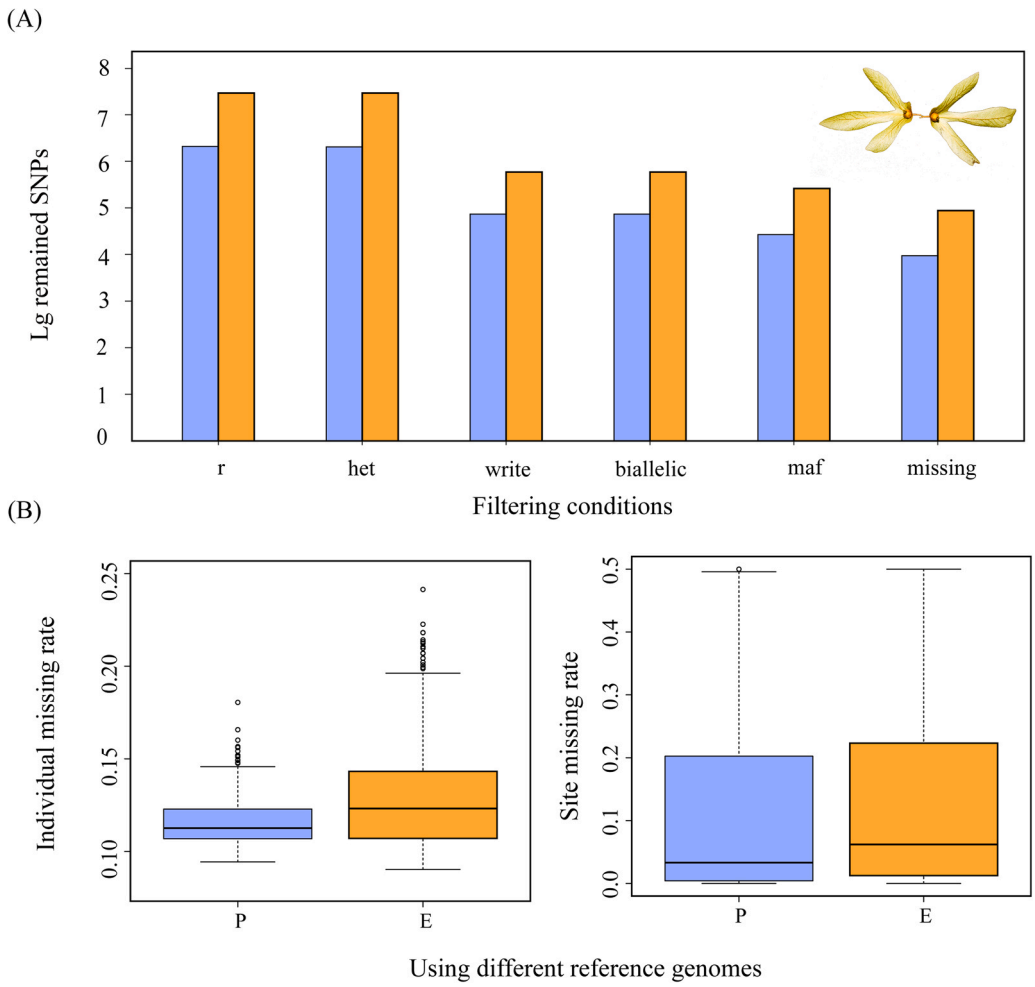

Fig. 2. The result of the number of retained SNPs, and the percentage of missing sites and individuals after all filtering steps. (A) The lg-transformed values of the number of SNPs from the initial to the final through filtering steps for different reference genomes (The blue color represents *P. stenoptera* as a reference genome, the orange color represents *E. roxburghiana* as the reference genome; r: –min-samples-per-pop 0.6; het: –max-obs-het 0.7; write: –write-single-snp; biallelic: –biallelic-only; maf: –maf 0.05; missing: –max-missing 0.5). (B) Boxplots displaying the percentage of missing sites and individual through the different reference genome (P: *P. stenoptera* as the reference genome; E: *E. roxburghiana* as the reference genome).

indexed the reference genomes, mapped the sequence reads to them, created loci by incorporating paired-end reads, and subsequently applied various filtering options (Fig. 1).

After SNP calling and filtering (Supplementary Information: Experimental Procedures), our study suggested that there are large differences in the number of SNPs when employing both reference-based strategies. Utilizing the species itself as the reference genome can generate high-quality SNPs that are an order of magnitude larger than that obtained by using a closely related one (Fig. 2A; Table S4). Notably, when comparing the individual and sites missing rates of the final SNPs obtained through both methods, no significant differences were detected (Fig. 2B). In both scenarios, utilizing the species itself as the reference genome, in which results emerged as the optimal choice. SNPs contain a wealth of information, and they have proven to be one of the most abundant forms of genetic variation between individuals of a species (Ghosh et al., 2002). The huge discrepancy in the number of SNPs might suggest that a more substantial volume of valid, high-quality information can be acquired using this species as a reference genome, because it can help to obtain more accurate and detailed downstream analysis results. Meanwhile, this discovery also indicated that utilizing closely related ones, despite results in fewer SNPs, the quality remains satisfactory, making it a feasible option for numerous studies.

A crucial conclusion of our research revealed a significant difference in SNP acquisition between both reference-based approaches, shedding light on the impact of sequence divergence in the reference genome. Compared to previous study that explored how the choice of reference genome affects the output of a bioinformatics pipeline (e.g., Bohling, 2020), our study uniquely emphasizes the significance of employing the species itself as a reference genome and advocate for whole genome sequencing when the species lacks a specific reference genome. Although employing a closely related species as a reference genome fulfills the requirements for numerous studies (e.g., Paris et al., 2017; Shen et al., 2023), the number of SNPs obtained by the closely related one is significantly reduced compared to methods that used the species itself as a reference genome (Fig. 2; Table S4). Furthermore, the kinship of the reference genome to the subject of study should also be taken into consideration. Using reference genomes from more distantly related species has resulted in unrealistically low transition/transversion (Ts/Tv) ratios, suggesting an increase rate of miscalling (Shafer et al., 2017). Therefore, given the decreasing cost of whole genome sequencing, our suggestion is to directly conduct whole genome sequencing for species lacking a specific reference genome, particularly those with smaller genomes. This approach allows for the acquisition of more valuable information and optimizes subsequent analyses. If a reference genome of a closely related species must be considered, our recommendation is to choose a genome that shares the closest relationship with the subject in studies. This approach minimizes miscalling issues and tends to yield satisfactory results.

In short, our results obtained from the diverse strategies applying in STACKS showed that selected different reference-based approaches have a significant influence on the number of high-quality SNPs. After the comparison of both approaches, we derived a set of recommendations for RAD-seq analysis using STACKS. Employing the species itself as the reference genome emerges as the optimal choice for SNP calling using STACKS in analyses (Fig. 2; Table S4). Alternatively, when this approach is not feasible, using closely related ones are a good choice, while distantly related genomes should be avoided. The obvious discrepancy in numbers of SNPs following our conception from different reference genomes, really determines the different results. However, this study did not specifically explore the divergence between the two reference genomes due to differences in their assembly levels. The reference genome of *E. roxburghiana* is at the chromosome level, while the reference genome of *P. stenoptera* is at the scaffold level. Further investigations on this aspect will help us better understand how structure variants in the two genomes impact the number of SNPs acquired.

Funding

This work was supported by the National Natural Science Foundation of China (No. 42171063); Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (No. Y4ZK111B01); the Special Fund for Scientific Research of Shanghai Landscaping & City Appearance Administrative Bureau (G212406, G242414, G242416); the "Yunnan Revitalization Talent Support Program" in Yunnan Province (XDYC-QNRC-2022-0028); the CAS "Light of West China" Program; and the 14th Five-Year Plan of Xishuangbanna Tropical Botanical Garden, Chinese Academy Sciences (XTBG-1450303).

CRediT authorship contribution statement

Yi-Gang Song: Writing – review & editing, Supervision, Conceptualization. **Hong-Hu Meng:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization. **Jie Li:** Writing – review & editing, Conceptualization. **Tian-Rui Wang:** Writing – original draft, Software, Methodology, Formal analysis, Data curation. **Pei-Han Huang:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation. **Ou-Yan Fang:** Writing – review & editing. **Min Li:** Writing – review & editing. **Ren-Ping Su:** Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgement

We would like to thank the special issue "*Tree Biology: From Genomics to Genetic Improvement*" of Plant Science for inviting us to make contribution. Also, the three reviewers are grateful for the valuable suggestions and comments that help to improve the quality of this study.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.plantsci.2024.112109.

References

K.R. Andrews, J.M. Good, M.R. Miller, G. Luikart, R.A. Hohenlohe, Harnessing the power of RADseq for ecological and evolutionary genomics, Nat. Rev. Genet. 17 (2) (2016) 81–92, https://doi.org/10.1038/nrg.2015.28.

N.A. Baird, P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W. A. Cresko, E.A. Johnson, Rapid SNP discovery and genetic mapping using sequenced RAD markers, PLoS One 3 (10) (2008) e3376, https://doi.org/10.1371/journal.pone.0003376.

J. Bohling, Evaluating the effect of reference genome divergence on the analysis of empirical RADseq datasets, Ecol. Evol. 10 (14) (2020) 7585–7601, https://doi.org/10.1002/ece3.6483.

A. Casanova, F. Maroso, A. Blanco, M. Hermida, N. Ríos, G. García, A. Manuzzi, L. Zane, A. Verissimo, J.L. García-Marín, C. Bouza, M. Vera, P. Martínez, Low impact of different SNP panels from two building-loci pipelines on RAD-Seq population genomic metrics: case study on five diverse aquatic species, BMC Genom. 22 (1) (2021) 150, https://doi.org/10.1186/s12864-021-07465-w.

J.M. Catchen, A. Amores, P.A. Hohenlohe, W.A. Cresko, J.H. Postlethwait, Stacks: building and genotyping Loci de novo from short-read sequences, G3-GENES Genom. GENET 1 (2011) 171–182, https://doi.org/10.1534/g3.111.000240.

J.M. Catchen, P.A. Hohenlohe, S. Bassham, A. Amores, W.A. Cresko, Stacks: an analysis tool set for population genomics, Mol. Ecol. 22 (2013) 3124–3140, https://doi.org/10.1111/mec.12354.

J.W. Davey, M.L. Blaxter, RADSeq: next-generation population genetics, Brief. Funct. Genom. 9 (5) (2010) 416–423, https://doi.org/10.1093/bfgp/elr007.

J.W. Davey, P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, M.L. Blaxter, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, Nat. Rev. Genet. 12 (7) (2011) 499–510, https://doi.org/10.1038/nrg3012.

Y.M. Ding, X.X. Pang, Y. Cao, W.P. Zhang, S.S. Renner, D.Y. Zhang, W.N. Bai, Genome structure-based Juglandaceae phylogenies contradict alignment-based phylogenies and substitution rates vary with DNA repair genes, Nat. Commun. 14 (1) (2023) 617, https://doi.org/10.1038/s41467-023-36247-z.

H. Dittberner, C. Becker, W. Jiao, K. Schneeberger, N. Hölzel, A. Tellier, J.D. Meaux, Strengths and potential pitfalls of hay-transfer for ecological restoration revealed by RAD-seq analysis in two floodplain Arabis species, Mol. Ecol. 28 (2018) 3887–3901, https://doi.org/10.1101/474858.

B.H. Duan, S.M. Mu, Y.Q. Guan, S.Q. Li, Y. Yu, W.B. Liu, Z.J. Li, X.D. Ji, X.J. Kang, Genetic diversity and population structure of the swimming crab (*Portunus trituberculatus*) in China seas determined by genotyping-by-sequencing (GBS), Aquaculture 555 (2022) 738233, https://doi.org/10.1016/j.aquaculture.2022.738233.

D.A.R. Eaton, I. Overcast, ipyrad: Interactive assembly and analysis of RADseq datasets, Bioinformatics 36 (8) (2020) 2592–2594, https://doi.org/10.1093/bioinformatics/btz966.

R.J. Elshire, J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, S.E. Mitchell, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, PloS One 6 (5) (2011) e19379, https://doi.org/10.1371/journal.pone.0019379.

P.D. Etter, J.L. Preston, S. Bassham, W.A. Cresko, E.A. Johnson, Local de novo assembly of RAD paired-end contigs using short sequencing reads, PLoS One 6 (2011) e18561, https://doi.org/10.1371/journal.pone.0018561.

S. Ghosh, P. Malhotra, P.V. Lalitha, S. Guha-Mukherjee, V.S. Chauhan, Novel genetic mapping tools in plants: SNPs and LD-based approaches, Plant Sci. 162 (3) (2002) 329–333, https://doi.org/10.1016/S0168-9452(01)00587-8.

R.A. Gibbs, The Human Genome Project changed everything, Nat. Rev. Genet. 21 (10) (2020) 575–576, https://doi.org/10.1038/s41576-020-0275-3.

Y.M. Hsu, S.S. Wang, Y.C. Tseng, S.R. Lee, H. Fang, W.C. Hung, H.I. Kuo, H.Y. Dai, Assessment of genetic diversity and SNP marker development within peanut germplasm in Taiwan by RAD-seq, Sci. Rep. 12 (1) (2022) 14495, https://doi.org/10.1038/s41598-022-18737-0.

M.E. Hudson, Sequencing breakthroughs for genomic ecology and evolutionary biology, Mol. Ecol. Resour. 8 (1) (2008) 3–17, https://doi.org/10.1111/j.1471-8286.2007.02019.x.

M.H. Kang, H.L. Wu, H.H. Liu, W.Y. Liu, M.J. Zhu, Y. Han, W. Liu, C.L. Chen, Y. Song, L. N. Tan, K.Q. Yin, Y.S. Zhao, Z. Yan, S.L. Lou, Y.J. Zan, J.Q. Liu, The pan-genome and local adaptation of *Arabidopsis thaliana*, Nat. Commun. 14 (1) (2023) 6259, https://doi.org/10.1038/s41467-023-42029-4.

A.M. Lu, D.E. Stone, L.J. Grauke, Juglandaceae, Vol 4, in: Z.Y. Wu, P.H. Raven, D. Y. Hong (Eds.), Flora of China, Science Press & St. Louis: Missouri Botanical Garden Press, Beijing, 1999, pp. 278–280. Vol 4.

S.H. Martin, K.K. Dasmahapatra, N.J. Nadeau, C. Salazar, J.R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, C.D. Jiggins, Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies, Genome Res. 23 (11) (2013) 1817–1828, https://doi.org/10.1101/gr.159426.113.

S. McGinn, I.G. Gut, DNA sequencing - spanning the generations, N. Biotechnol. 30 (4) (2013) 366–372, https://doi.org/10.1016/j.nbt.2012.11.012.

H.H. Meng, C.Y. Zhang, S.L. Low, L. LI, J.Y. Shen, Nurainas, Y. Zhang, P.H. Huang, S. S. Zhou, Y.H. Tan, J. Li, Two new species from Sulawesi and Borneo facilitate phylogeny and taxonomic revision of *Engelhardia* (Juglandaceae), Plant Divers 44 (6) (2022b) 552–564, https://doi.org/10.1016/j.pld.2022.08.003.

H.H. Meng, C.Y. Zhang, Y.G. Song, X.Q. Yu, G.L. Cao, L. Li, C.N. Cai, J.H. Xiao, S.S. Zhou, Y.H. Tan, J. Li, Opening a door to the spatiotemporal history of plants from the tropical Indochina Peninsula to subtropical China, Mol. Phylogenet. Evol. 171 (2022a) 107458, https://doi.org/10.1016/j.ympev.2022.107458.

X.Y. Mu, L. Tong, M. Sun, Y.X. Zhu, J. Wen, Q.W. Lin, B. Liu, Phylogeny and divergence time estimation of the walnut family (Juglandaceae) based on nuclear RAD-Seq and

chloroplast genome data, Mol. Phylogenet. Evol. 147 (6) () (2020) 106802, https://doi.org/10.1016/J.YMPEV.2020.106802.
R. Orita, Y. Nagano, Y. Kawamura, K. Kimura, G. Kobayashi, Genetic diversity and population structure of razor clam *Sinonovacula constricta* in Ariake Bay, Japan, revealed using RAD-Seq SNP markers, Sci. Rep. 11 (1) (2021) 7761, https://doi.org/10.1038/s41598-021-87395-5.
E. Pante, J. Abdelkrim, A. Viricel, D. Gey, S.C. France, M.C. Boisselier, S. Samadi, Use of RAD sequencing for delimiting species, Heredity 114 (5) (2015) 450–459, https://doi.org/10.1038/hdy.2014.105.
J.R. Paris, J.R. Stevens, J.M. Catchen, Lost in parameter space: a road map for stacks, Methods Ecol. Evol. 8 (10) (2017) 1360–1373, https://doi.org/10.1111/2041-210X.12775.
B.K. Peterson, J.N. Weber, E.H. Kay, H.S. Fisher, H.E. Hoekstra, Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species, PloS One 7 (5) (2012) e37135, https://doi.org/10.1371/journal.pone.0037135.
M. Piwczyński, L. Granjon, P. Trzeciak, J. Carlos Brito, M.O. Popa, M.D. Dinka, N. P. Johnston, Z. Boratyński, Unraveling phylogenetic relationships and species boundaries in the arid adapted *Gerbillus* rodents (Muridae: Gerbillinae) by RAD-seq data, Mol. Phylogenet. Evol. 189 (2023) 107913, https://doi.org/10.1016/j.ympev.2023.107913.
W. Probowati, S. Koga, K. Harada, Y. Nagano, A.J. Nagano, K. Ishimaru, K. Ohshima, S. Fukuda, RAD-Seq analysis of wild Japanese garlic (*Allium macrostemon* Bunge) growing in Japan revealed that this neglected crop was previously actively utilized, Sci. Rep. 13 (1) (2023) 16354, https://doi.org/10.1038/s41598-023-43537-5.
N.C. Rochette, J.M. Catchen, Deriving genotypes from RAD-seq short-read data using Stacks, Nat. Protoc. 12 (2017) 2640–2659, https://doi.org/10.1038/nprot.2017.123.
B.E. Rubin, R.H. Ree, C.S. Moreau, Inferring phylogenies from RAD sequence data, PloS One 7 (4) (2012) e33394, https://doi.org/10.1371/journal.pone.0033394.
K. Ruegg, E.C. Anderson, J. Boone, J. Pouls, T.B. Smith, A role for migration-linked genes and genomic islands in divergence of a songbird, Mol. Ecol. 23 (19) (2014) 4757–4769, https://doi.org/10.1111/mec.12842.
A.B. Shafer, C.R. Peart, S. Tusso, I. Maayan, A. Brelsford, C.W. Wheat, J.B. Wolf, Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference, Methods Ecol. Evol. 8 (2017) 907–917, https://doi.org/10.1111/2041-210X.12700.
Y.T. Shen, G. Yao, Y.F. Li, X.L. Tian, S.M. Li, N. Wang, C.J. Zhang, F. Wang, Y.P. Ma, RAD-seq data reveals robust phylogeny and morphological evolutionary history of the ornamentally important plant genus, *Rhododendron*, Hortic. Plant J. (2023), https://doi.org/10.1016/j.hpj.2022.11.010.
D.E. Stone, Review of new world *Alfaroa* and old world *Alfaropsis* (Juglandaceae), Novon 20 (2) (2010) 215–224, https//doi.org/10.3417/2009027.
N. Su, R.G.J. Hodel, X. Wang, J.R. Wang, S.Y. Xie, C.X. Gui, L. Zhang, Z.Y. Chang, L. Zhao, D. Potter, J. Wen, Molecular phylogeny and inflorescence evolution of *Prunus* (Rosaceae) based on RAD-seq and genome skimming analyses, Plant Divers 45 (4) (2023) 397–408, https://doi.org/10.1016/j.pld.2023.03.013.
D. Torkamaneh, J. Laroche, F. Belzile, Genome-Wide SNP Calling from Genotyping by Sequencing (GBS) Data: A Comparison of Seven Pipelines and Two Sequencing Technologies, PloS One 11 (8) (2016) e0161333, https://doi.org/10.1371/journal.pone.0161333.
M. Uhlen, S.R. Quake, Sequential sequencing by synthesis and the next-generation sequencing revolution, Trends Biotechnol. 41 (12) (2023) 1565–1572, https://doi.org/10.1016/j.tibtech.2023.06.007.
C.E. Wagner, I. Keller, S. Wittwer, O.M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, O. Seehausen, Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation, Mol. Ecol. 22 (3) (2013) 787–798, https://doi.org/10.1111/mec.12023.
T. Wang, S.Y. Duan, C. Xu, Y. Wang, X.Z. Zhang, X.F. Xu, L.Y. Chen, Z.H. Han, Z.T. Wu, Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits, Nat. Commun. 14 (1) (2023) 7377, https://doi.org/10.1038/s41467-023-43270-7.
S. Wang, E. Meyer, J.K. McKay, M.V. Matz, 2b-RAD: a simple and flexible method for genome-wide genotyping, Nat. Methods 9 (2012) 808–810, https://doi.org/10.1038/nmeth.2023.
C.Y. Zhang, S.L. Low, Y.G. Song, Nurainas, G. Kozlowski, T.V. Do, L. Li, S.S. Zhou, Y. H. Tan, G.L. Cao, Z. Zhou, H.H. Meng, J. Li, Shining a light on species delimitation in the tree genus *Engelhardia* Leschenault ex Blume (Juglandaceae), Mol. Phylogenet. Evol. 152 (2020) 106918, https://doi.org/10.1016/j.ympev.2020.106918.
Y.J. Zhao, G.S. Yin, X. Gong, RAD-sequencing improves the genetic characterization of a threatened tree peony (*Paeonia ludlowii*) endemic to China: Implications for conservation, Plant Divers 45 (5) (2023) 513–522, https://doi.org/10.1016/j.pld.2022.07.002.