

# Detecting clonemate pairs in multicellular diploid clonal species based on a shared heterozygosity index

Lei Yu<sup>1</sup>  | John J. Stachowicz<sup>2,3</sup> | Katie DuBois<sup>2,3</sup> | Thorsten B. H. Reusch<sup>1</sup>

<sup>1</sup>Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

<sup>2</sup>Center for Population Biology, University of California, Davis, California, USA

<sup>3</sup>Department of Evolution and Ecology, University of California, Davis, California, USA

## Correspondence

Lei Yu, Marine Evolutionary Ecology, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.  
Email: [lyu@geomar.de](mailto:lyu@geomar.de)

## Funding information

China Scholarship Council, Grant/Award Number: 201704910807; Human Frontiers of Science Program, Grant/Award Number: RGP0042\_2020

**Handling Editor:** Frederic Austerlitz

## Abstract

Clonal reproduction, the formation of nearly identical individuals via mitosis in the absence of genetic recombination, is a very common reproductive mode across plants, fungi and animals. To detect clonal genetic structure, genetic similarity indices based on shared alleles are widely used, such as the Jaccard index, or identity by state. Here we propose a new pairwise genetic similarity index, the SH index, based on segregating genetic marker loci (typically single nucleotide polymorphisms) that are identically heterozygous for pairs of samples ( $N_{SH}$ ). To test our method, we analyse two old seagrass clones (*Posidonia australis*, estimated to be around 8500 years old; *Zostera marina*, >750 years old) along with two young *Z. marina* clones of known age (17 years old). We show that focusing on shared heterozygosity amplifies the power to distinguish sample pairs belonging to different clones compared to methods focusing on all shared alleles. Our proposed workflow can successfully detect clonemates at a location dominated by a single clone. When the collected samples involve two or more clones, the SH index shows a clear gap between clonemate pairs and interclone sample pairs. Ideally  $N_{SH}$  should be on the order of approximately  $\geq 3000$ , a number easily achievable via restriction-site associated DNA (RAD) sequencing or whole-genome resequencing. Another potential application of the SH index is to detect possible parent–descendant pairs under selfing. Our proposed workflow takes advantage of the availability of the larger number of genetic markers in the genomic era, and improves the ability to distinguish clonemates from nonclonemates in multicellular diploid clonal species.

## KEYWORDS

clonal reproduction, clone, clonemate, genet, ramet, seagrass

## 1 | INTRODUCTION

Clonal reproduction of multicellular species refers to the formation of new individuals that are genetically nearly identical to the parent via mitosis (De Meeüs et al., 2007). This reproductive mode occurs

without genetic recombination and is widespread among the plant, fungal and animal kingdoms (Buss, 1983; Honnay & Bossuyt, 2005). It comes in many different forms, such as separation of cells from the parent through fission, budding or branching, but also processes involving single cell stages such as parthenogenesis (animals),

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

apomixis (plants), and gynogenesis or pseudogamy (De Meeûs et al., 2007; Orive & Krueger-Hadfield, 2021). In clonal species, a module or ramet (Harper & White, 1974) is the basic modular unit with an ability to live independently, such as an individual shoot in grasses or a single polyp in corals. The product of a single zygote from gamete fusion (Noble et al., 1979) collectively forms a genet (synonymous with clone), which comprises a population of ramets (here also clonemates) that emerge via mitosis during its lifetime. The majority of clonal species can conduct both clonal and sexual reproduction (Schön et al., 2009; Stoeckel, Arnaud-Haond, & Krueger-Hadfield, 2021; Stoeckel, Porro, & Arnaud-Haond, 2021), and thus have two levels of biological organization. The asexual population of ramets belonging to a single clone/genet is nested into the population of clones/genets (Reusch et al., 2021), making the proper detection of the clonemates vs. nonclonemates an important starting point to study the ecology or evolution of such species.

As clonemates are often phenotypically undistinguishable, they can no longer be physically tracked once they obtain autonomy and break off the parental module, which is the case in most clonal species including seagrasses, corals, algae or fungi. Hence, genetic markers such as microsatellites or AFLPs have been widely used for detecting clonemates (Arnaud-Haond et al., 2007). Genotyping with many codominant molecular markers such as microsatellites has been particularly useful in the past, as any given multilocus genotype (MLG) can be assigned a likelihood to emerge by chance (i.e., through sexual reproduction and recombination). However, identical or nearly identical MLGs are not definitive evidence for clonal reproduction (Halkett et al., 2005), if for example there is low (or no) allelic diversity at particular loci. To calculate the probability of finding identical MLGs resulting from distinct zygotes, the population frequencies of segregating alleles are required, and the population is assumed to be under Hardy–Weinberg equilibrium (Parks & Werth, 1993). Estimates of such population-level frequencies are inaccurate in populations dominated by a few clones (Arnaud-Haond et al., 2007), and meaningless if a site is dominated by one clone (Arnaud-Haond et al., 2012; Edgeloe et al., 2022; Japaud et al., 2015; Reusch et al., 1999; Smith et al., 1992).

An alternative suite of approaches is the analysis of a pairwise genetic distance or similarity matrix, which can identify clonemates with slightly different MLGs, and also works with dominant/recessive markers such as AFLPs (Duhovnikoff & Dodd, 2003). The genomes of clonemates are identical by descent (IBD), except at sites with novel somatic mutations, as the genomes are inherited from the same asexual ancestor via mitosis without recombination. Because somatic mutations are rare, clonemates are predicted to reveal a smaller genetic distance (or higher levels of genetic similarity) than ramets from different clones. This will lead to a bimodal pattern in the frequency distribution of pairwise genetic distances/similarities. A threshold can then be defined to separate the clonemates vs. inter-clone ramet pairs (Bailleul et al., 2016; Duhovnikoff & Dodd, 2003; Meirmans, 2020). The currently available genetic similarity indices (e.g., Jaccard index, identity by state) are based on shared alleles. However, when genetic diversity is low in a population, shared

similarity may be high and the distinction in similarity is blurred between clonemates and nonclonemates. Here, we propose a novel pairwise genetic similarity index, the shared heterozygosity (SH) index, which describes the number of shared heterozygous sites divided by the number of heterozygous sites observed in the individual with the higher count. The SH index includes only heterozygous loci under the logic that heterozygous sites are the most informative ones for detecting clonemates because they show the potential for segregation in sexual reproduction, while remaining identically heterozygous in clonal reproduction.

We applied our workflow to two of the oldest seagrass clones in marine systems detected so far, a *Zostera marina* clone in Finland (>750 years old, Yu et al., 2020) and a *Posidonia australis* clone in Australia (estimated to be around 8500 years old, Edgeloe et al., 2022), which probably represented the extreme end of accumulating somatically generated variation. We also applied the workflow to two 17-year-old *Z. marina* clones cultured in outdoor flow-through mesocosms, representing the recent accumulating somatically generated variation. We examined whether our method could reliably detect clonemates based on seagrass ramets belonging to one, two and multiple (more than two) clones, respectively.

## 2 | MATERIALS AND METHODS

### 2.1 | Introducing the SH index for pairs of samples

For two samples ( $X_1$  and  $X_2$ ) in multicellular diploid clonal species,  $N_{SH}$  denotes the number of the genetic markers where two samples are identically heterozygous;  $N_{Het}(X_n)$  denotes the number of the heterozygous loci in  $X_n$ .

$$SH(X_1) = N_{SH} / N_{Het}(X_1)$$

$$SH(X_2) = N_{SH} / N_{Het}(X_2)$$

$$SH = \min(SH(X_1), SH(X_2))$$

When technical replicates are available, the clonemate pairs can be identified based on the closeness of their SH relative to technical replicates. If technical replicates are unavailable, a threshold needs to be set using minimal representations of intraclone SH values (Figure 1).

### 2.2 | Detecting clonemates based on a published data set on seagrass *P. australis*

Edgeloe et al. (2022) reported a *P. australis* clone spanning at least 180 km, which was estimated to be around 8500 years old based on vegetative expansion rate. Although the clone was polyploid, the single nucleotide polymorphisms (SNPs) were called assuming the clone was diploid. The SNP data set containing 18,021 SNPs for 133 individuals (including four technical replicates) from 10 populations

## 1. DNA extraction

Generating technical replicates

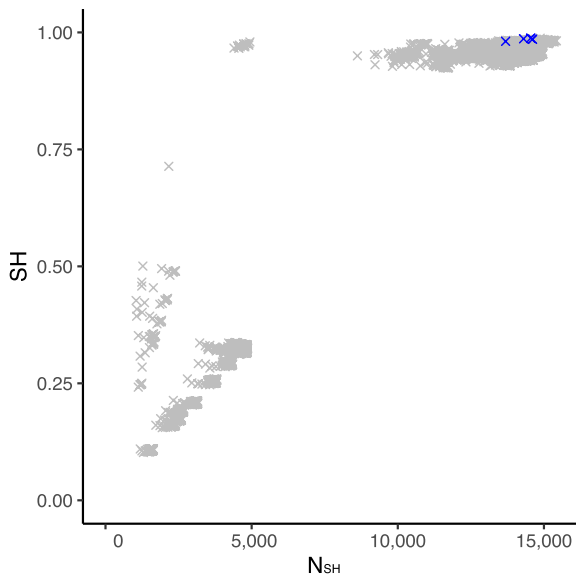
## 2. Genotyping

(i.e., multiple pieces of tissue from the same sample are used independently for all the processes from DNA extraction to genotyping).

## 3. SH and $N_{SH}$

For each pair of samples, calculate  $N_{SH}$  and SH based on the loci where both samples have available genotype calls (i.e., not missing data).

### 4.1 with technical replicates



### 4.2 without technical replicates

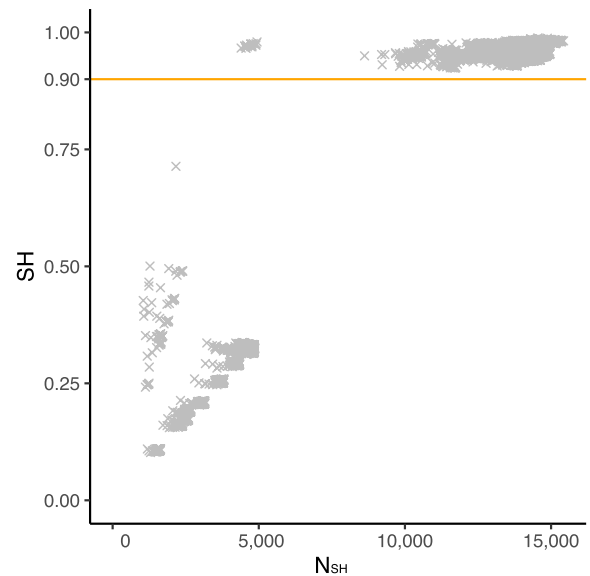


**FIGURE 1** Workflow for detecting clonemate pairs in multicellular diploid clonal species based on the SH index. Sample pairs representing technical replicates are marked in blue. When technical replicates are available, the sample pairs with the SH index close to those of the technical replicates represent clonemate pairs. If unavailable, a threshold needs to be set to separate the intra- and interclone sample pairs, and the mode close to  $SH = 1$  represents clonemate pairs. The figures presented here are based on the data in Figure 5.

was used here (named as “all.vcf” in the Dryad repository <https://doi.org/10.5061/dryad.0cfxpnw4p>). We calculated the SH index along with  $N_{SH}$  for each pair of samples based on the nucleotide positions in the reference genome where both samples provided genotype calls using a Python script (<https://github.com/leiyu37/Detecting-clonemates.git>).

### 2.3 | Jaccard similarity index based on shared alleles vs. shared heterozygosity using the data set on seagrass *P. australis*

We assessed the advantage of focusing on SH by comparing a Jaccard similarity index calculated in two different ways ([https://github.com/leiyu37/Detecting-clonemates/tree/main/res/Jaccard\\_similarity](https://github.com/leiyu37/Detecting-clonemates/tree/main/res/Jaccard_similarity)). We first calculated the Jaccard similarity index

based on the collection of all the alleles present in each sample across all the loci. To assess the advantage of focusing on shared heterozygosity, we also calculated the index based on the collection of all the heterozygous genomic positions in each sample across all the loci.

### 2.4 | Detecting clonemates based on two *Z. marina* clones of known age (17 years)

Two *Z. marina* clones (a subsample of eight in total) were kept for 17 years in large, 300-L outdoor flow-through mesocosms at Bodega Marine Laboratory (BML) under ambient light and temperature conditions (Hughes et al., 2009). Each genet was initiated with a single terminal shoot collected from Bodega Harbour, California, in July 2004. Six shoots (= ramets) were collected from each clone for

genomic analysis in 2021, referred to as the "GREEN" and "RED" clone, respectively.

Genomic DNA was extracted using a NucleoSpin plant II kit from Macherey-Nagel following the manufacturer's instructions. The plant tissue (160–200 mg fresh weight) was ground together with liquid nitrogen. Library construction and Illumina sequencing (PE150, paired-end 150 bp) were conducted by BGI (Beijing Genomics Institute). The quality of the raw Illumina reads was assessed by FASTQC version 0.11.5 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). BBDOUK (last modified November 7, 2019, <https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/>) was used to remove adapter contamination and conduct a basic quality filtering: (i) reads with more than one N were discarded (maxns = 1); (ii) reads shorter than 50 bp after trimming were discarded (minlength = 50); (iii) reads with average quality below 10 after trimming were discarded (maq = 10). FASTQC version 0.11.5 was used for a second round of quality check for the filtered reads.

The quality-filtered Illumina reads were mapped against the chromosome-level *Z. marina* reference genome version 3.1 (Ma et al., 2021) using BWA MEM version 0.7.17 (Li & Durbin, 2009). The alignments were converted to BAM format and then sorted using SAMTOOLS version 1.11 (Li et al., 2009). The MarkDuplicates module in GATK version 4.1.1.0 (Van der Auwera & O'Connor, 2020) was used to identify and tag duplicate reads in the BAM files. Sequencing depth and percentage of covered genome region for each ramet were calculated using SAMTOOLS version 1.11 and BEDTOOLS version 2.25.0 (Quinlan & Hall, 2010), respectively. HaplotypeCaller (GATK version 4.1.1.0) was used to generate a GVCF format file for each ramet, and all the GVCF files were combined by CombineGVCFs (GATK version 4.1.1.0). GenotypeGVCFs (GATK version 4.1.1.0) was used to call genetic variants.

BCFTOOLS version 1.11 (Li, 2011) was used to remove SNPs within 20 bp of an indel or other variant type. We then extracted only bi-allelic SNPs located in the six main chromosomes (2,969,820 SNPs). INFO annotations were extracted using VariantsToTable (GATK version 4.1.1.0). SNPs meeting one or more than one of the following criteria were marked by VariantFiltration (GATK version 4.1.1.0):  $MQ < 40.0$ ;  $FS > 60.0$ ;  $QD < 13.0$ ;  $MQRandSum > 2.5$  or  $MQRandSum < -2.5$ ;  $ReadPosRandSum < -2.5$ ;  $ReadPosRandSum > 2.5$ ;  $SOR > 3.0$ ;  $DP > 1192.45$  ( $2 \times$  average DP). Those SNPs were excluded by SelectVariants (GATK version 4.1.1.0). A total of 1,105,628 SNPs were retained. VCFTOOLS version 0.1.15 (Danecek et al., 2011) was used to convert individual genotypes to missing data when  $GQ < 30$  or  $DP < 20$ . Individual homozygous reference calls with one or more than one read supporting the variant allele, and individual homozygous variant calls with one or more than one read supporting the reference allele, were also converted to missing data using a custom Python script (<https://github.com/leiyu37/Detecting-clonemates.git>). Finally, we kept only genomic sites showing polymorphism (604,066 SNPs). The SH index and  $N_{SH}$  were calculated for each pair of samples.

## 2.5 | Detecting clonemates based on a published data set on seagrass *Z. marina*

Yu et al. (2020) conducted whole-genome resequencing of 24 ramets of a *Z. marina* clone located in Finland, hereafter referred to as the Finnish *Z. marina* clone. Based on growth and areal expansion data, the Finnish *Z. marina* clone was estimated to be between 750 and 1500 years old. All 24 ramets were sequenced by Illumina HiSeq, while three of them were additionally sequenced independently by Illumina Novaseq to a very high coverage (1300 $\times$ ), providing technical replicates. A subset was selected from the Novaseq data ( $\sim 80\times$ ) to achieve a similar coverage to that of the HiSeq data. We then re-analysed the HiSeq and Novaseq ( $\sim 80\times$ ) data together using the same method as above for the two 17-year-old clones. All the SNP calling and filtering processes were identical to those for the 17-year-old clones, except the threshold for DP during hard filtering ( $DP > 2945.57$ ), because the average DP increased with the sample size. The SH index and  $N_{SH}$  were calculated for each pair of samples.

We assessed the influence of  $N_{SH}$  on the SH index by recalculating SH using 1/10, 1/100 and 1/1000 of the 38,831 SNPs for the 24 ramets reported in Yu et al. (2020), and the corresponding  $N_{SH}$  was around 3000, 300 and 30, respectively. Each selection was repeated 100 times, and calculation was conducted for each pair of samples for each of the 100 repeats.  $N_{SH}$  and the SH index from the 100 repeats were pooled for the subsequent analysis.

## 3 | RESULTS

### 3.1 | Jaccard similarity index based on shared alleles vs. shared heterozygosity using the data set on seagrass *P. australis*

When calculated based on shared alleles, the mean value of the Jaccard similarity index for the clonemate pairs and interclone sample pairs was 0.97 and 0.64, respectively (Figure 2a). As for the calculation based on shared heterozygosity, the mean value of the Jaccard index remained similar for the clonemate pairs (i.e., 0.94), but decreased to 0.26 for interclone sample pairs (Figure 2b). This comparison showed that focusing on shared heterozygosity could amplify the difference between intra- and interclone modes.

### 3.2 | Detecting clonemates based on the data set of the Finnish *Z. marina* clone

After re-analysing the data (HiSeq data for the 24 ramets, and Novaseq data for three of the ramets), we obtained 41,885 SNPs, and the pairwise  $N_{SH}$  ranged from 28,619 to 32,210. The HiSeq and Novaseq data for the three ramets provided three pairs of technical replicates, and the minimum SH between technical replicates was 0.9998. The SH index between clonemates were distributed within a

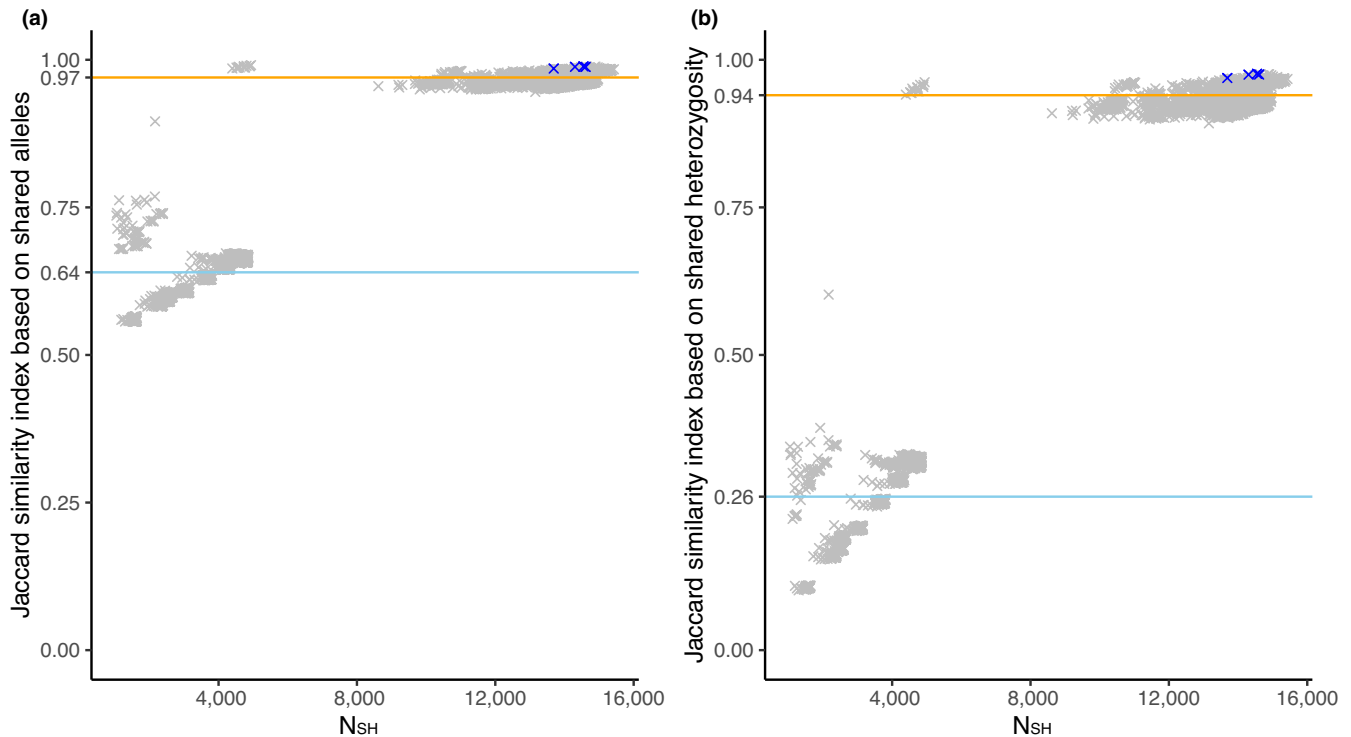


**FIGURE 2** Jaccard similarity index based on a published data set on seagrass *Posidonia australis*. Pairs of technical replicates are marked in blue. The orange and skyblue horizontal lines indicate the mean value of the Jaccard index for clonemate pairs and interclone sample pairs, respectively. (a) Jaccard similarity index based on shared alleles. For each sample pair, Jaccard similarity index is calculated based on the collection of all the alleles present in each sample across all the loci. (b) Jaccard similarity index based on shared heterozygosity. For each sample pair, Jaccard similarity index is calculated based on the collection of all the heterozygous genomic positions in each sample across all the loci. The scripts for the calculation are available on GitHub (<https://github.com/leiyu37/Detecting-clonemates.git>).

narrow range from 0.9548 to 0.9993, close but clearly distinct from the values for technical replicates. Our analysis confirmed that all 24 ramets belonged to the same clone (Figure 3).

### 3.3 | Influence of $N_{SH}$ on the SH index

By calculating the SH index and  $N_{SH}$  based on the subsets of the SNP data set in Yu et al. (2020), we found that smaller  $N_{SH}$  led to larger variation of the SH index (Figure 4). The standard deviation was 0.0076, 0.0107 and 0.0259 for subsets “1/10,” “1/100” and “1/1000,” respectively. The consistently high SH-value when  $N_{SH}$  was at least on the order of  $\geq 3000$  suggested this number of  $N_{SH}$  as a reasonable minimum target to apply this method.

### 3.4 | Detecting clonemates based on the data set of seagrass *P. australis*

Based on the SNP data set encompassing 18,021 SNPs for 133 individuals (including four technical replicates) from 10 populations, the pairwise  $N_{SH}$  ranged from 1047 to 15,428. The minimum SH was 0.9809 and 0.9242 for pairs of technical replicates and clonemates, respectively. The SH index for clonemate pairs were close to those

of the technical replicates (Figure 5). The maximum SH for ramet pairs from different clones was 0.7138 ( $N_{SH} = 2165$ ), while all the other ramet pairs from different clones showed  $SH < 0.51$ . When  $N_{SH} \geq 3000$ , the difference in terms of SH index between clonemate pairs and interclone ramet pairs was  $> 0.41$ , which showed as a wide gap between intra- and interclone modes. According to our results, population GU included one clone comprising six ramets, while the ramets from the other populations belonged to a different clone, confirming the clone assignments in Edgeloe et al. (2022).

### 3.5 | Detecting clonemates based on the two 17-year-old seagrass *Z. marina* clones

We then applied our method to a situation of two *Z. marina* clones of known age (17 years old) where technical replicates were unavailable. The whole-genome resequencing data of the 12 cultured *Z. marina* ramets covered  $> 93\%$  of the reference genome, and reached  $> 66\times$  coverage (Table 1). Pairwise  $N_{SH}$  ranged from 108,585 to 288,606. The SH index showed a wide gap between the intra- and interclone modes, and the mode with SH close to 1 represented clonemate pairs (Figure 6). Within the clone “GREEN,” the minimum pairwise SH was 0.9975. We detected one “contaminant” ramet (i.e., R10) in the “RED” clone. The five ramet pairs composed of R10 and

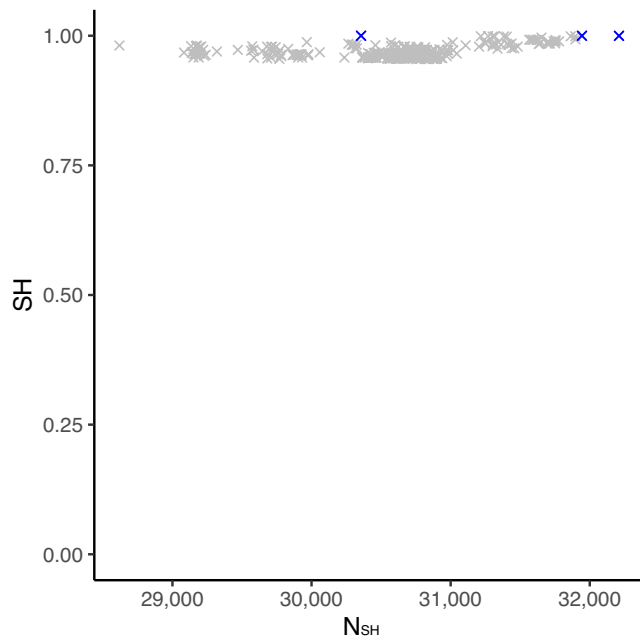


**FIGURE 3** Detecting clonemate pairs using the SH index and technical replicates based on a published data set on seagrass *Zostera marina*. Each data point represents one pair of samples. Sample pairs representing technical replicates are marked in blue. The SH index values of all the sample pairs are close to those of the technical replicates, meaning that all samples belong to the same clone.

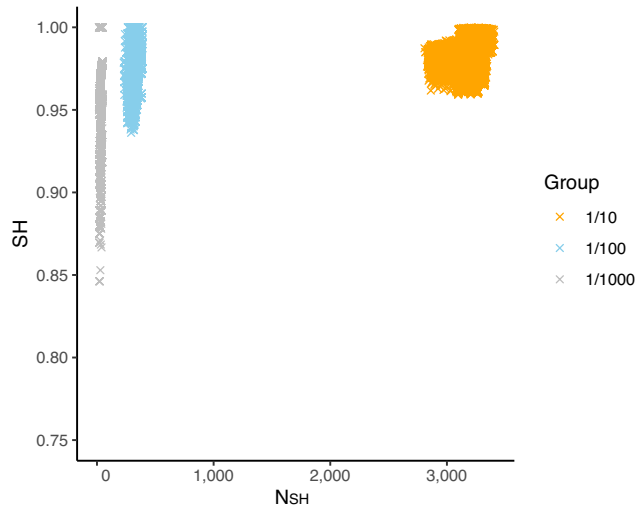


**FIGURE 4** Distribution of SH index values within the Finnish *Zostera marina* clone based on subsets (1/10, 1/100 and 1/1000) of SNPs. The variation of SH increases when  $N_{SH}$  decreases. The standard deviation is 0.0076, 0.0107 and 0.0259 for subsets “1/10,” “1/100” and “1/1000,” respectively.

one of the other five “RED” ramets showed SH ranging from 0.4103 to 0.4114, indicating that R10 was not a member of this clone. With the exception of ramet R10, the minimum SH for the ramet pairs among the other five “RED” ramets was 0.9974. The interclone SH ranged from 0.4790 to 0.4803.

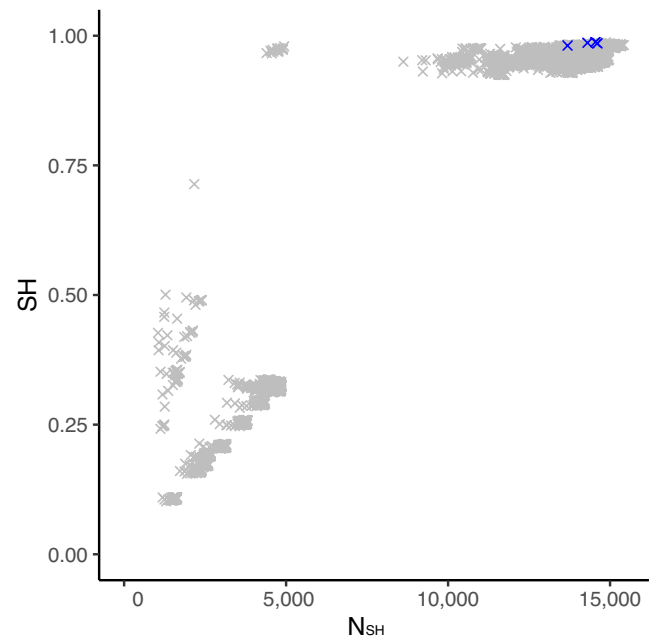


**FIGURE 5** Detecting clonemate pairs using the SH index and technical replicates based on a published data set on the seagrass *Posidonia australis*. Each data point represents one sample pair. The data set contains four pairs of technical replicates, which are marked in blue. The sample pairs with the SH index close to those of technical replicates represent clonemate pairs.

**TABLE 1** Coverage of the whole-genome resequencing data for the two cultured 17-year-old *Zostera marina* clones (“RED” and “GREEN”)

| Sample | Average depth for the covered region | Genome region covered (%) |
|--------|--------------------------------------|---------------------------|
| G01    | 68.64                                | 93.33                     |
| G02    | 80.42                                | 93.66                     |
| G06    | 79.05                                | 93.72                     |
| G07    | 69.94                                | 93.36                     |
| G10    | 77.49                                | 93.52                     |
| G16    | 73.31                                | 93.26                     |
| R02    | 81.55                                | 93.86                     |
| R03    | 85.13                                | 94.06                     |
| R05    | 68.14                                | 93.57                     |
| R08    | 66.44                                | 93.51                     |
| R09    | 71.73                                | 93.59                     |
| R10    | 67.97                                | 93.69                     |

## 4 | DISCUSSION

In many multicellular clonal species, the degree of clonality is highly variable (for seagrasses, e.g., Reusch & Boström, 2011; for corals, e.g., Miller & Ayre, 2004), including the extreme case of a single site dominated by one clonal lineage. Examples of an entire location covered by one clone include fungi (Smith et al., 1992), the seagrass

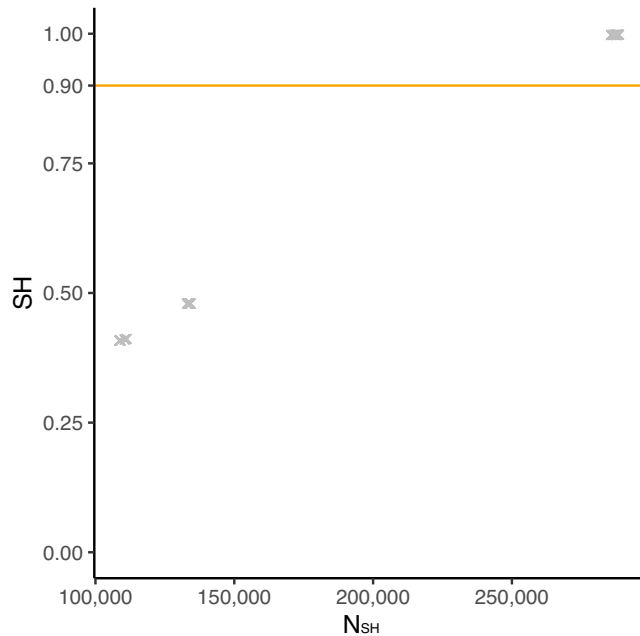


**FIGURE 6** Detecting clonemate pairs using the SH index without technical replicates based on the two 17-year-old *Zostera marina* clones. A threshold needs to be set to distinguish between intra- and interclone modes, and the mode close to  $SH = 1$  represents clonemate pairs.

species *Z. marina* (Reusch et al., 1999; Yu et al., 2020), *Thalassia testudinum* (Bricker et al., 2018), *Posidonia oceanica* (Arnaud-Haond et al., 2012) and *P. australis* (Edgeloe et al., 2022), and the coral genus *Acropora* (Japaud et al., 2015). The situation of having a population in which all sampled entities belong to the same genet is thus more widespread than previously thought. MLG-based methods cannot work in locations dominated by a few or a single clone, due to the difficulty in calculating the likelihood of an MLG based on population allele frequencies which are principally unavailable under such circumstances. An alternative way for clonemate detection is to use an index to assess the similarity of the collected samples. Ideally, such an index demonstrates clear differences between the intra- and interclone sample pairs. In our study, focusing on shared heterozygosity amplified the power to distinguish sample pairs belonging to different clones than methods focusing on all shared alleles (Figure 2). Our SH index is similar to the Jaccard index calculated based on shared heterozygosity, with the difference being that the denominator for the calculation of the SH index is the size of the larger set, instead of the size of the union of the two sets in the Jaccard Index. However, the key here is to quantify genetic similarity based on shared heterozygosity, while the exact index can be different. As a result, the SH index is more suitable for detecting clonemates in locations dominated by a few or a single clone, and can generally improve the performance of the methods based on a pairwise genetic similarity matrix.

The SH index can potentially be used to detect possible parent–descendant pairs under selfing (Reusch, 2001), because this reproductive mode leads to loss of heterozygosity in every round of sexual

reproduction. A case in point is the analysis of the 17-year-old *Z. marina* clones where we found one “contaminant” ramet (i.e., R10) that clearly did not belong to the clone “RED.” One possible reason is that R10 was produced by sexual event between members of the clone “RED” (i.e., selfing). In the first 15 years of growth, this would be the only way for contamination to occur, as the clones were cultured in isolated containers. Under selfing, each of the heterozygous loci in the parent can either remain identically heterozygous or change to a homozygous state in R10. As a consequence, the heterozygous loci in R10 should be a subset of those in the “RED” clone, and thus  $SH(R10)$  (see Introduction for the SH index method) for the five ramet pairs composed of R10 and one of the other five “RED” ramets should be close to 1. However, that is not the case, as  $SH(R10)$  for the five ramet pairs ranged from 0.4320 to 0.4331. Hence, we can exclude the possibility that R10 is derived from selfing of the “RED” clone.

Application of the SH index requires a sufficient number of genetic markers where the two target samples are identically heterozygous (i.e.,  $N_{SH}$ ). With the continuing development of sequencing technology and decreasing costs, it is now straightforward to obtain thousands of genetic markers using whole-genome resequencing, restriction-site associated DNA (RAD) sequencing or SNP arrays. However, the level of  $N_{SH}$  can still be low based on genome-wide marker sets, when the target species shows very low levels of genome-wide heterozygosity. Prior to further analyses, it is advisable to verify whether or not  $N_{SH}$  is high enough (approximately  $N_{SH}$  on the order of  $\geq 3000$ ). In conclusion, our workflow takes advantage of the availability of the larger number of genetic markers in the genomic era, and improves the identification of clonemates in multicellular diploid clonal species.

## AUTHOR CONTRIBUTIONS

L.Y. and T.B.H.R. conceived and designed the study, J.J.S. and K.D. cultured and provided the samples of the two 17-year-old seagrass *Z. marina* clones, L.Y. extracted the DNA, analysed the sequencing data, and conducted the bioinformatic analyses, and L.Y. and T.B.H.R. wrote the paper, with input from all authors.

## ACKNOWLEDGEMENTS

This study was supported by a 4-year PhD scholarship from the China Scholarship Council (CSC) to L.Y. (No. 201704910807), and by a Human Frontiers of Science (HFSP) grant to T.B.H.R. (RGP0042\_2020). We thank D. Gill for laboratory assistance, and Dr Y. Li for the comments on an early version of the manuscript. Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

DNA sequence data are available in the NCBI short read archive, BioProject no. PRJNA806459, SRA accession nos. SRR18000159–SRR18000170. Custom-made computer code is available at Github: <https://github.com/leiyu37/Detecting-clonemates>.

## SAMPLING PERMIT

The original sampling of the founder ramets for the two 17-year-old *Z. marina* clones was conducted under permit no. SC-4874 to John Stachowicz, while the Finish *Z. marina* ramets were sampled under permit no. MH 5448/2015 to Christoffer Boström.

## ORCID

Lei Yu  <https://orcid.org/0000-0003-3748-5470>

## REFERENCES

- Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular Ecology*, 16(24), 5115–5139. <https://doi.org/10.1111/j.1365-294X.2007.03535.x>
- Arnaud-Haond, S., Duarte, C. M., Diaz-Almela, E., Marbà, N., Sintes, T., & Serrão, E. A. (2012). Implications of extreme life span in clonal organisms: Millenary clones in meadows of the threatened seagrass *Posidonia oceanica*. *PLoS One*, 7(2), e30454. <https://doi.org/10.1371/journal.pone.0030454>
- Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: A package to identify MultiLocus clonal lineages and handle clonal data sets in R. *Methods in Ecology and Evolution*, 7(8), 966–970. <https://doi.org/10.1111/2041-210X.12550>
- Bricker, E., Calladine, A., Virnstein, R., & Waycott, M. (2018). Mega clonality in an aquatic plant—A potential survival strategy in a changing environment. *Frontiers in Plant Science*, 9, 435. <https://doi.org/10.3389/fpls.2018.00435>
- Buss, L. W. (1983). Evolution, development, and the units of selection. *Proceedings of the National Academy of Sciences of the United States of America*, 80(5), 1387–1391. <https://doi.org/10.1073/pnas.80.5.1387>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De Meeüs, T., Prugnolle, F., & Agnew, P. (2007). Asexual reproduction: Genetics and evolutionary aspects. *Cellular and Molecular Life Sciences*, 64(11), 1355–1372. <https://doi.org/10.1007/s00018-007-6515-2>
- Douhovnikoff, V., & Dodd, R. S. (2003). Intra-clonal variation and a similarity threshold for identification of clones: Application to *Salix exigua* using AFLP molecular markers. *Theoretical and Applied Genetics*, 106(7), 1307–1315. <https://doi.org/10.1007/s00122-003-1200-9>
- Edgeloe, J. M., Severn-Ellis, A. A., Bayer, P. E., Mehravi, S., Breed, M. F., Krauss, S. L., Batley, J., Kendrick, G. A., & Sinclair, E. A. (2022). Extensive polyploid clonality was a successful strategy for seagrass to expand into a newly submerged environment. *Proceedings of the Royal Society B: Biological Sciences*, 289(1976), 20220538. <https://doi.org/10.1098/rspb.2022.0538>
- Halkett, F., Simon, J., & Balloux, F. (2005). Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, 20(4), 194–201. <https://doi.org/10.1016/j.tree.2005.01.001>
- Harper, J. L., & White, J. (1974). The demography of plants. *Annual Review of Ecology and Systematics*, 5(1), 419–463. <https://doi.org/10.1146/annurev.es.05.110174.002223>
- Honnay, O., & Bossuyt, B. (2005). Prolonged clonal growth: Escape route or route to extinction? *Oikos*, 108(2), 427–432. <https://doi.org/10.1111/j.0030-1299.2005.13569.x>
- Hughes, R. A., Stachowicz, J. J., & Williams, S. L. (2009). Morphological and physiological variation among seagrass (*Zostera marina*) genotypes. *Oecologia*, 159(4), 725–733. <https://doi.org/10.1007/s00442-008-1251-3>
- Japaud, A., Bouchon, C., Manceau, J., & Fauvelot, C. (2015). High clonality in *Acropora palmata* and *Acropora cervicornis* populations of Guadeloupe, French Lesser Antilles. *Marine and Freshwater Research*, 66(9), 847–851. <https://doi.org/10.1071/MF14181>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Ma, X., Olsen, J. L., Reusch, T. B. H., Procaccini, G., Kudrna, D., Williams, M., Grimwood, J., Rajasekar, S., Jenkins, J., Schmutz, J., & van de Peer, Y. (2021). Improved chromosome-level genome assembly and annotation of the seagrass, *Zostera marina* (eelgrass). *F1000research*, 10, 1–13. <https://doi.org/10.12688/f1000research.38156.1>
- Meirmans, P. G. (2020). Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, 20(4), 1126–1131. <https://doi.org/10.1111/1755-0998.13145>
- Miller, K. J., & Ayre, D. J. (2004). The role of sexual and asexual reproduction in structuring high latitude populations of the reef coral *Pocillopora damicornis*. *Heredity (Edinb)*, 92(6), 557–568. <https://doi.org/10.1038/sj.hdy.6800459>
- Noble, J. C., Bell, A. D., & Harper, J. L. (1979). The population biology of plants with clonal growth: I. the morphology and structural demography of *Carex arenaria*. *Journal of Ecology*, 67, 983–1008. <https://doi.org/10.2307/2259224>
- Orive, M. E., & Krueger-Hadfield, S. A. (2021). Sex and Asex: A clonal lexicon. *The Journal of Heredity*, 112(1), 1–8. <https://doi.org/10.1093/jhered/esaa058>
- Parks, J. C., & Werth, C. R. (1993). A study of spatial features of clones in a population of bracken fern, *Pteridium aquilinum* (Dennstaedtiaceae). *American Journal of Botany*, 80(5), 537–544. <https://doi.org/10.1002/j.1537-2197.1993.tb13837.x>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Reusch, T. B. H. (2001). Fitness-consequences of geitonogamous selfing in a clonal marine angiosperm (*Zostera marina*). *Journal of Evolutionary Biology*, 14(1), 129–138. <https://doi.org/10.1046/j.1420-9101.2001.00257.x>
- Reusch, T. B. H., Baums, I. B., & Werner, B. (2021). Evolution via somatic genetic variation in modular species. *Trends in Ecology & Evolution*, 36(12), 1083–1092. <https://doi.org/10.1016/j.tree.2021.08.011>
- Reusch, T. B. H., & Boström, C. (2011). Widespread genetic mosaicism in the marine angiosperm *Zostera marina* is correlated with clonal reproduction. *Evolutionary Ecology*, 25(4), 899–913. <https://doi.org/10.1007/s10682-010-9436-8>
- Reusch, T. B. H., Boström, C., Stam, W. T., & Olsen, J. L. (1999). An ancient eelgrass clone in the Baltic. *Marine Ecology Progress Series*, 183, 301–304. <https://doi.org/10.3354/meps183301>
- Schön, I., Martens, K., & van Dijk, P. (2009). *Lost sex: The evolutionary biology of parthenogenesis*. Springer Dordrecht.
- Smith, M. L., Bruhn, J. N., & Anderson, J. B. (1992). The fungus *armillaria bulbosa* is among the largest and oldest living organisms. *Nature*, 356(6368), 428–431. <https://doi.org/10.1038/356428a0>
- Stoeckel, S., Arnaud-Haond, S., & Krueger-Hadfield, S. A. (2021). The combined effect of haplodiplonty and partial clonality on genotypic and genetic diversity in a finite mutating population. *The Journal of Heredity*, 112(1), 78–91. <https://doi.org/10.1093/jhered/esaa062>

- Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2021). The discernible and hidden effects of clonality on the genotypic and genetic states of populations: Improving our estimation of clonal rates. *Molecular Ecology Resources*, 21(4), 1068–1084. <https://doi.org/10.1111/1755-0998.13316>
- Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using docker, GATK, and WDL in Terra*. O'Reilly Media.
- Yu, L., Boström, C., Franzenburg, S., Bayer, T., Dagan, T., & Reusch, T. B. H. (2020). Somatic genetic drift and multilevel selection in a clonal seagrass. *Nature Ecology and Evolution*, 4(7), 952–962. <https://doi.org/10.1038/s41559-020-1196-4>

**How to cite this article:** Yu, L., Stachowicz, J. J., DuBois, K., & Reusch, T. B. H. (2023). Detecting clonemate pairs in multicellular diploid clonal species based on a shared heterozygosity index. *Molecular Ecology Resources*, 23, 592–600. <https://doi.org/10.1111/1755-0998.13736>