Archimer
http://archimer.ifremer.fr

Microsatellite markers developed through pyrosequencing allow clonal discrimination in the invasive alga *Caulerpa taxifolia*

Sophie Arnaud-Haond[1, *], Rui Candeias[2], Ester A. Serrão[2], Sara JL Teixeira[2],

[1] Ifremer, Department RBE, UMR 212 - Ecosystème Marin Exploité (EME) Bd Jean Monnet, BP 171, 34203, Sète Cedex, France
[2] CCMAR, CIMAR-Laboratório Associado, Faculdade de Ciências e Tecnologias, Universidade do Algarve, Gambelas, 8005-139, Faro, Portugal

*: Corresponding author : Sophie Arnaud-Haond, email address : sarnaud@ualg.pt

Abstract:

Polymorphic microsatellites were developed for the invasive green algae *Caulerpa taxifolia* using next generation DNA sequencing. Results showed a limited rate of microsatellites for the amount of sequences, possibly explaining failure of previous attempts for microsatellite development through classical methods. Eight polymorphic loci were selected that exhibited polymorphism and a null or negligible rate of amplification failure. The number of alleles per locus ranged from two to seven. The reconstruction of Multi Locus Genotypes and the heterozygosity and departure from Hardy–Weinberg equilibrium confirmed the influence of clonal reproduction and showed the usefulness of this set of markers to successfully discriminate clonal lineages and analyze the clonal and genetic composition of algal beds. These markers will be used to further investigate the clonal composition and genetic structure in populations of *C. taxifolia* and to attempt retracing the origin of and pathways followed by invasive clonal lineages.

Keywords: Caulerpa taxifolia ; Algae ; Invasion ; Mediterranean ; Queensland ; New South Wales ; Seagrass

The tropical green alga *Caulerpa taxifolia*, native of north-eastern Australia, was first recorded in the Mediterranean Sea in 1984 (Meinesz & Hesse 1991) and in the early 2000"s in Southern Australia (Schaffelke *et al.* 2002; Millar 2004), where it has since colonized widespread areas forming dense beds with important detrimental effects including the displacement of many native species (Wright et al. 2007; Boudouresque et al. 2009). Molecular tools can help retracing the origin and pathways of invasions. For clonal species as *C. taxifolia* (Smith & Walters 1999) discriminating clonal lineages on the basis of their Multi Locus Genotype (MLG, Arnaud-Haond et al. 2005) is a prerequisite to understand the demography and connectivity of populations. Previous failure in developing microsatellites on *Caulerpa* species were mostly due to difficulties in obtaining good quality and high yield DNA extracts (Varela-Alvarez et al. 2006) to build enriched banks. Molecular studies were therefore mostly based on phylogeography of ITS supporting a unique event of accidental aquarium release (Meinesz & Boudouresque 1996; Jousson *et al.* 1998) as the origin of the Mediterranean invasion, but the low polymorphism did not elucidate the recent introduction in New South Wales (NSW, Southern Australia; Schaffelke et al. 2002). A first attempt to use Amplified Fragment Length Polymorphism (Murphy & Schaffelke 2003) showed enhanced resolution but suggested that new markers with more polymorphism would allow a more precise appraisal of the origin, pathways and mechanisms of invasion.

Two trials of next generation „454" sequencing were performed by Biocant Park, Portugal, with a 1/8 genome coverage per sample. Samples were whole genomic DNA isolated from three individuals of *C. taxifolia* preserved at -80°C, extracted using the CTAB method (Doyle & Doyle 1987). These generated a total of 218,435 reads of which 1110 contained di- to hexa-nucleotide microsatellites (about 5‰) with a minimum of 8 repeats identified using MSATCOMMANDER (Faircloth 2008). Primers could be designed with PRIMER3 (Rozen & Skalesty 2000) for only 43 microsatellites (only 3.9% of sequences had a length that permitted the definition of primers in flanking regions).

From a total of 29 primer pairs tested, 9 (GenBank accession n°JX458334 to JX458342) revealed polymorphism on a first subset of 15 sampling units from several localities. These were used to genotype two sampling sets from the native area in Queensland (25 and 19 sampling-units collected respectively in Townsville, 2010, and Brisbane, 1999). No consistent amplification could be obtained on 15 sampling-units of the congeneric species *C. racemosa* and *C. prolifera*. Due to a high rate of failure in amplification of one locus on samples of Townsville (20% failure, no missing data in Brisbane) only eight loci were eventually retained (Table 1).

Amplifications in 10 µL reactions contained 0.4 ng of genomic DNA, 1x Go*Taq* Flexi buffer, 1.5 µM MgCl$_2$ (except GE3, 2 µM), 200 µM of dNTP"s, 0.3 µM of each primer and 0.1 U of Go*Taq* polymerase (Promega). PCR amplifications were conducted on a Perkin-Elmer Gene Amp System 7200 (Waltham, MA, USA) following conditions: 10 min. at 94°C; 35 cycles composed of 40s of denaturation at 94°C, 40s at the annealing temperature (Table 1) and 40s elongation at 72°C and a final 30 min. elongation step at 72°C. Fragments were separated on an ABI 3130 XL automatic sequencer (Applied Biosystems, Foster City, CA, USA) with the internal size standard Rox 350. Alleles were scored using Peak Scanner version 1.0 (Applied Biosystems).

Data were analyzed with GenClone (Arnaud-Haond & Belkhir 2007). Allelic richness varied from 3 to 7 alleles.locus-1 (Table 1). Replicated MLG were found in Townsville (24 distinct MLG for 25 sampling-units) and Brisbane (11 distinct MLG for 19 sampling-units). The probability that identical MLG originated independently through distinct events of sexual recombination (estimated using both P$_{sex}$(F_{IS}) and P$_{sex}$) was very low in all cases (p<0.01), allowing identical MLG to be confidently assigned as clonal replicates (i.e., the same clonal lineage). This confirmed the importance of clonal propagation in the dynamics of *C. taxifolia* (Ceccherelli & Cinelli 1999; Smith & Walters 1999), particularly in Brisbane. Datasets with

one copy of each MLG (i.e., after removal of clonal replicates) were characterized for heterozygosity and F_{IS} using the package Genetix (Belkhir et al. 1996-2001), showing high variation among loci and sampling set, with significant and sometimes extreme departure from Hardy-Weinberg equilibrium, a common feature in clonal organisms (Halkett et al. 2005). The occurrence of null alleles is an unlikely explanation considering the low rate of amplification failure (<2% sampling-unit.locus^{-1} in samples from Townsville, none in Brisbane), and the inconsistency in the pattern of F_{IS}. The limited scale of sampling (1 km^2) renders the hypothesis of a spatial Whalund effect unlikely to explain the heterozygote deficiency observed for most loci in Kissing Point. Since technical explanations are not supported by evidence, departures from HWE are more likely to have a biological origin. The simplest explanation is the influence of clonal reproduction, which can cause high variance in F_{IS} such as the one observed here (Halkett et al. 2005) and enhance inbreeding and selfing opportunity in this monoecious species (Phillips 2009). Also, the recent recolonization of Kissing Point after the exceptional disappearance of *Caulerpa* beds during the wet season 2009-2010, may have resulted in some mosaic genetic patchiness if recruited algae originated from several, differentiated sources dispersing distant clonal lineages.

Results obtained here show replicated MLG and several features characteristic of a significant influence of clonal versus sexual reproduction on the genetic composition (heterozygosity and HWE) of algal beds. These markers should be useful to study the clonal and genetic diversity in natural beds of *C. taxifolia*, and to attempt identifying clonal lineages and retracing the origin and pathways of the invasive ones in different regions of the world.

Acknowledgements

References

Arnaud-Haond S, Alberto F, Teixeira S, Procaccini G, Serrao EA, Duarte CM (2005) Assessing genetic diversity in clonal organisms: Low diversity or low resolution? Combining power and cost efficiency in selecting markers. Journal of Heredity, 96, 434-440.

Belkhir K, Borsa P, Chikhi L, N.Raufaste, Bonhomme F (1996-2001) GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome et Populations, Interactions, Adaptations, CNRS UMR5000, Université Montpellier II, Montpellier (France).

Boudouresque CF, Bernard G, Pergent G, Shili A, Verlaque M (2009) Regression of Mediterranean seagrasses caused by natural processes and anthropogenic disturbances and stress: a critical review. Botanica Marina, 52, 395-418.

Ceccherelli G, Cinelli F (1999) The role of vegetative fragmentation in dispersal of the invasive alga *Caulerpa taxifolia* in the Mediterranean. Marine Ecology-Progress Series, 182, 299-303.

Doyle JJ, Doyle JLI (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin, 11, 11-15.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources, 8, 92-94.

Halkett F, Simon JC, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. Trends in Ecology & Evolution, 20, 194-201.

Jousson O, Pawlowski J, Zaninetti L, Meinesz A, Boudouresque CF (1998) Molecular evidence for the aquarium origin of the green alga *Caulerpa taxifolia* introduced to the Mediterranean Sea. Marine Ecology-Progress Series, 172, 275-280.

Meinesz A, Boudouresque CF (1996) On the origin of *Caulerpa taxifolia* in the Mediterranean Sea. Comptes Rendus De L Academie Des Sciences Serie Iii-Sciences De La Vie-Life Sciences, 319, 603-613.

Meinesz A, Hesse B (1991) Introduction of the Tropical Alga *Caulerpa-Taxifolia* and Its Invasion of the Northwestern Mediterranean. Oceanologica Acta, 14, 415-426.

Millar AJK (2004) New records of marine benthic algae from New South Wales, eastern Australia. Phycological Research, 52, 117-128.

Murphy NE, Schaffelke B (2003) Use of amplified fragment length polymorphism (AFLP) as a new tool to explore the invasive green alga *Caulerpa taxifolia* in Australia. Marine Ecology-Progress Series, 246, 307-310.

Phillips JA (2009) Reproductive ecology of *Caulerpa taxifolia* (Caulerpaceae, Bryopsidales) in subtropical eastern Australia. European Journal of Phycology, 44, 81-88.

Rozen S, Skalesty HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In:Krawetz S MS (ed) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa. pp. 365-386.

Schaffelke B, Murphy N, Uthicke S (2002) Using genetic techniques to investigate the sources of the invasive alga *Caulerpa taxifolia* in three new locations in Australia. Marine Pollution Bulletin, 44, 204-210.

Smith CM, Walters LJ (1999) Fragmentation as a strategy for Caulerpa species: Fates of fragments and implications for management of an invasive weed. Marine Ecology-Pubblicazioni Della Stazione Zoologica Di Napoli I, 20, 307-319.

Varela-Alvarez E, Andreakis N, Lago-Leston A, Pearson GA, Serrao EA (2006) Genomic DNA isolation from green and brown algae (Caulerpales and Fucales) for microsatellite library construction. Journal of Phycology, 42, 741-745.

Wright JT, McKenzie LA, Gribben PE (2007) A decline in the abundance and condition of a native bivalve associated with *Caulerpa taxifolia* invasion. Mar Freshwater Res, 58, 263-272.

Table 1. Characterization of eight polymorphic microsatellite loci identified in *Caulerpa taxifolia* including genbank accession numbers, motif, primer sequence and annealing temperature (Ta ºC) for each locus. Subsequent columns detail allele size range (in bp), allelic richness (A), observed (H_{obs}) and expected (H_{exp}) heterozygosity, and departure from Hardy Weinberg Equilibrium (F_{IS}, in bold when significant after a 1000 permutation test) observed in the MLG detected in the native areas of Townsville (N=22, top values in each row) and Brisbane (N=11, lower values in each row).

Locus	GenBank	Repeat motif	Primer sequence (5'-3')	Ta (°C)	Range	A	H_{obs}	H_{exp}	F_{IS}
GE3	JX458334	(AG)9	F: ACGTACTGTGTATACTGAAAGGG R: CAAGGTACTGTTGCCCGTC	58	220-228	5	0.57	0.71	**0.20**
					220-228	4	0.64	0.69	**-0.15**
GE15	JX458335	(ACCT)13	F: CTCGGAGAGTCCGTAAGCC R: TTTCCCGGTAGTAACTTGGG	54	250-294	5	0.17	0.30	**0.46**
					268-272	2	0.27	0.25	**-0.20**
GE16	JX458336	(AAGG)11	F: CCCTCACGAAGCGCAAAC R: CGTTGCGGGTTCCATGTC	56	152-164	4	0.38	0.62	**0.40**
					164-168	2	1.0	0.52	**-1**
GE21	JX458338	(GGTTT)17	F: GGTCTTGGTCTAGCGCCTC R: TACGTGCTTCTCCGGTTCG	56	282-314	2	0.00	0.09	**1**
					324	1	0	0	-
GX6	JX458339	(AAGG)17	F: ACCCGTAGGCCAACCAAG R: ACCTTCGGTGCTACCTCAC	54	186-224	6	0.17	0.57	**0.72**
					220-224	2	0.09	0.09	-0.03
GX7	JX458340	(AGCCTT)18	F: CCTAGGCCACAACAACTGG R: CCGAAAGTGTTCACGGAGG	54	230-244	4	0.38	0.52	**0.28**
					336	1	0	0	-
GX8	JX458341	(AAGCC)20	F: TCGCCTTGGTGAAGGTCTG R: CTCGGCGGAGTTCCTAGTG	54	155-171	4	0.67	0.59	**-0.14**
					166-171	2	0.45	0.37	**-0.25**
GX9	JX458342	(AGCCTT)18	F: TCCTAGGCCACAACTAACTGG R: GGAGGTGGTTAAACATGACCC	54	125-157	4	0.04	0.50	**0.92**
					224	1	0	0	-