Phylogeographic differentiation versus transcriptomic adaptation to warm temperatures in Zostera marina, a globally important seagrass

A. JUETERBOCK,* S. U. FRANSSEN,†‡ N. BERGMANN,§ J. GU,† J. A. COYER,¶ T. B. H. REUSCH,†† E. BORNBERG-BAUER‡ and J. L. OLSEN**

*Faculty of Biosciences and Aquaculture, Nord University, Universitetsallén 11, Bodø 8049, Norway, †Institut für Populationsgenetik, Vetmeduni Vienna, Veterinärplatz 1, Vienna 1210, Austria, ‡Institute for Evolution and Biodiversity, University of Münster, Hüfferstr. 1, Münster 48149, Germany, §Integrated School of Ocean Sciences (ISOS), Kiel University, Leibnizstr. 3, Kiel 24098, Germany, ¶Shoals Marine Laboratory, University of New Hampshire, Durham, NH 03824, USA, **Groningen Institute for Evolutionary Life Sciences, Ecological and Evolutionary Genomics Group, University of Groningen, P.O. Box 11103, Groningen 9700 CC, The Netherlands, ††GEOMAR Helmholtz-Centre for Ocean Research Kiel, Evolutionary Ecology of Marine Fishes, Düsternbrookweg 20, Kiel 24105, Germany

Abstract

Populations distributed across a broad thermal cline are instrumental in addressing adaptation to increasing temperatures under global warming. Using a space-for-time substitution design, we tested for parallel adaptation to warm temperatures along two independent thermal clines in Zostera marina, the most widely distributed seagrass in the temperate Northern Hemisphere. A North–South pair of populations was sampled along the European and North American coasts and exposed to a simulated heatwave in a common-garden mesocosm. Transcriptomic responses under control, heat stress and recovery were recorded in 99 RNAseq libraries with ~13,000 uniquely annotated, expressed genes. We corrected for phylogenetic differentiation among populations to discriminate neutral from adaptive differentiation. The two southern populations recovered faster from heat stress and showed parallel transcriptomic differentiation, as compared with northern populations. Among 2389 differentially expressed genes, 21 exceeded neutral expectations and were likely involved in parallel adaptation to warm temperatures. However, the strongest differentiation following phylogenetic correction was between the three Atlantic populations and the Mediterranean population with 128 of 4711 differentially expressed genes exceeding neutral expectations. Although adaptation to warm temperatures is expected to reduce sensitivity to heatwaves, the continued resistance of seagrass to further anthropogenic stresses may be impaired by heat-induced downregulation of genes related to photosynthesis, pathogen defence and stress tolerance.

Keywords: common-garden experiment, differential expression, global warming, heatwave, RNAseq, transcriptomics

Introduction

Seagrass ecosystems have experienced massive die-offs over the last decades due to increasing stresses including disease, invasive species, sediment and nutrient runoff, habitat loss through dredging and aquaculture, rising sea levels and global warming (Orth et al. 2006; Waycott et al. 2009). Heatwaves are predicted to become frequent in southern Europe and North America by 2100 (Easterling et al. 2000; Meehl et al. 2007), and are a major threat for Zostera marina (Reusch et al. 2006).
et al. 2008), the predominant seagrass in the Northern Hemisphere (Green & Short 2003; Olsen et al. 2004). For example, sustained temperatures of ≥25 °C during the summer of 2003 increased mortality and reduced shoot density by up to about 50% in a population in the Baltic Sea (Reusch et al. 2005); and on both the West (W) and East (E) Atlantic coasts (Chesapeake Bay, VA; Ria Formosa, PT; and the Brittany coast of France), summer temperatures now regularly reach temperatures of ≥25 °C (Nejrup & Pedersen 2008).

Understanding geographic variation in sensitivity to increasing heat stress can help to more realistically predict climate change-induced range shifts of Z. marina (Lavergne et al. 2010; Sinclair et al. 2010) and to identify thermally robust source populations for potential restoration (Procaccini et al. 2007). Common-garden experiments using populations from different geographical locations employ a space-for-time design to address potential adaptation to increasing heat stress but with the caveat that the end result reflects past evolutionary adaptation and thus cannot directly infer contemporary rates of adaptation (Kinnison & Hendry 2001; Reusch 2014).

Previous common-garden experiments with Z. marina revealed some evidence for local thermal adaptation of southern vs. northern populations (Franssen et al. 2011, 2014; Winters et al. 2011). Mediterranean populations are restricted to the thermally stable subtidal zone (Laugier et al. 1999) and, in contrast to Atlantic populations, have regularly experienced summer temperatures >26 °C over the past decade (Bergmann et al. 2010; Franssen et al. 2014). Accordingly, an Italian population (Adriatic Sea) performed better than Danish populations (Kattegat and the Baltic Sea) under common-garden experiments simulating the heatwave occurring in summer 2003. Individuals from the Italian population lost fewer shoots, were less responsive in osmoprotective metabolites (Gu et al. 2012) and more resilient in photosynthetic performance (Franssen et al. 2011; Winters et al. 2011; Gu et al. 2012). Such phenotypic divergence between northern and southern populations of Z. marina suggests reduced sensitivity to heatwaves at the species’ southern edge of its distribution.

However, phenotypic divergence may have been driven by both DNA-based changes and heritable epigenetic changes. Epigenetic variations are molecular-level changes that alter gene expression, but not the underlying DNA sequence, via histone modifications, chromatin remodelling, small interfering RNAs and DNA methylation (Bossdorf et al. 2008). In contrast to phenotypic variation within generations, including nonheritable physiological or behavioural responses, epigenetic variation may be heritable and persist even following long-term acclimation over generations (reviewed by Hirsch et al. 2012; Reusch 2014). Although this may be considered a shortcoming of common-garden studies, the inclusion of epigenetic carry-over effects may provide a more holistic picture of evolutionary potential in the context of rapid environmental change as compared with studies that only assess DNA-based changes (Richards et al. 2012; Zhang et al. 2013; Kilvitis et al. 2014).

Modification of gene expression can also drive adaptive evolution by linking molecular heritable changes at the DNA level with fitness-relevant traits (Emerson et al. 2010; Wittkopp 2013). Previous common-garden experiments revealed differences in the post-heatwave recovery patterns of gene expression (termed transcriptomic resilience) between Mediterranean and Danish populations (Franssen et al. 2011). While the observed differences were striking, the experimental design did not make it possible to determine whether the divergence was due to adaptive evolution, and if so, whether temperature was the major selective force as opposed to neutral processes or gene flow (reviewed in Merilä & Hendry 2014).

Methods to infer adaptive evolution of phenotypic differences include genotypic and phenotypic estimates of selection, comparison to models of neutral evolutionary change, reciprocal transplant experiments and \(Q_{ST} - F_{ST}\) comparisons (reviewed in Merilä & Hendry 2014). In \(Q_{ST} - F_{ST}\) comparisons, adaptive evolution is inferred when the phenotypic among-population divergence (\(Q_{ST}\)) exceeds among-population divergence at neutral genetic markers (\(F_{ST}\)) (reviewed in Leinonen et al. 2013). \(Q_{ST} - F_{ST}\) comparisons correct for phylogeographic differentiation and recently have been revised to a multivariate method (Leinonen et al. 2013) that more accurately discriminates neutral from adaptive divergence (Ovaskainen et al. 2011). Only when transcriptomic differences are correlated with temperature differences across replicate, independent locations can adaptive differentiation be attributable to temperature as the selective force (Kawecki & Ebert 2004; Merilä & Hendry 2014).

We refer to adaptive differentiation as only that portion of transcriptomic differentiation that exceeds neutral phylogeographic differentiation across multiple populations from contrasting thermal environments.

Here, we test the hypothesis that Z. marina shows adaptive differentiation in gene expression between thermally contrasting environments replicated on the North American and European coasts. More specifically: (i) Is there evidence for adaptation to temperature, as judged by heritable transcriptomic differentiation that exceeds neutral phylogenetic differentiation between the two southern and the two northern populations? (ii) Do southern populations show gene expression patterns consistent with reduced
sensitivity to heatwaves as evidenced by faster recovery from heat stress?

Methods

Sampling

Individuals of Zostera marina were sampled in April 2010 from northern (N) and southern (S) populations in Europe (Doverodde, NW Denmark 56°43.070′N 008°28.446′E, hereafter NE; Gabicce Mare, NE Italy 43°57.970′N 12°45.860′E, hereafter SE) and in the northeastern USA (Great Bay, NH 43°3.868′N, 70°52.345′W, hereafter NU; Waquoit, MA 41°33.240′N, 70°30.650′W, hereafter SU) (Fig. 1a). Note that the sampling site south USA (SU) does not represent the south of the USA but the southernmost of our US samples. The coastal region encompassing the N and S site of North America is characterized by one of the steepest latitudinal thermal gradients in the world’s oceans (fig. 2b in Frank et al. 2007; Wahle et al. 2013). Thus, even though the geographic distance between the North American sites is much less than between the N and S European sites, the differences in summer temperatures are comparable (Fig. 1b). Variability in water temperatures at the sampling locations was based on daily average sea surface temperature values recorded during summer months (June 1 to September 30 in years 2002–2011) over the past decade (Fig. 1b). Temperature data were extracted for sites NU, NE and SE from the NOAA \_OI\_SST\_V2 dataset (0.25° resolution, described in Reynolds et al. 2002), provided by NOAA/OAR/ESRL/PSC, Boulder, Colorado, USA, on http://www.esrl.noaa.gov/psd/). For site SU, that was not covered by the NOAA\_OI\_SST\_V2 dataset, we extracted temperature data from the National Estuarine Research Reserve System (http://cdmo.baruch.sc.edu/, station Sage Lot). Three to four clones with ≥3 shoots/clone were sampled from each of 10 patches at each location with a ≥5 m distance between samples to minimize chances of collecting the same clone (genotype) twice. Genotypic uniqueness of each experimental ramet was confirmed by genotyping the samples on an ABI 3100 Capillary Sequencer at four microsatellite loci (GenBank Accession nos.: AJ009898, AJ009900, AJ249305, AJ249307, Reusch 2000).

Experimental design

Within 48 h after collection, the plants were transported in seawater-filled cooling boxes to the AQUATRON, a mesocosm facility at the University of Münster, Germany. Details of the AQUATRON facility are described in Winters et al. (2011) and Fig. 1e,f. Briefly, each of two temperature-controlled water circuits supplied artificial seawater (31 psu) from a storage tank to six 700-L tanks (101 cm × 120 cm × 86.5 cm). Similar water chemistry between the two circuits was ensured with a water exchange rate of 1200 L/h. Each tank was populated with ~50 periwinkles (Littorina littorea) to regulate epiphytic growth. Each tank contained eight boxes (two boxes (37 cm × 27 cm) per population) with 10 genotypes. Shoots were planted in 10 cm of natural sediment (collected from Falckenstein, DE, in the Western Baltic Sea: 54°24.367′N, 010°11.438′E).

Plants were acclimated for 20 days to equilibrate temperature and light conditions (~400 μmol photons/s/m²) in order to minimize nonheritable differences in gene expression (Hoffmann et al. 2005; Whitehead & Crawford 2006). After 20 days, the temperature was raised 0.5 °C per day to 19 °C, the experimental control temperature in six of the 12 tanks over the entire experiment.

Heatwave simulation

After 20 days of acclimation at 19 °C, the temperature was raised in six of the 12 tanks at 1 °C per day to 25.5 °C, then held constant for 20 days to simulate the heatwave that occurred in the Baltic Sea during the summer of 2003 (Reusch et al. 2005). Finally, the temperature was decreased 1 °C per day to 19 °C and subsequently held for 20 days to allow the plants to recover (Fig. 1f).

RNA extraction

Samples for RNAseq (2-cm-long leaf tips) were excised from each ramet (three genotypes per treatment per population) at two time points under acute heat stress (T2 and T3: 0 and 5 days at 25.5 °C) and at three time points under recovery (T5, T7 and T9: 1, 20 and 30 days at 19 °C) (Fig. 1f, and Fig. S1, Supporting information). Tissue samples were immediately frozen in liquid nitrogen. RNA extraction was performed using the Invitrap Spin Plant RNA Mini Kit (Stratek Molecular) following the manufacturer’s protocol. We used the provided RP buffer for lysis. RNA concentrations and purity were tested by Nanodrop measurement (ND-1000, peQLab). RNA integrity was checked with an automated electrophoresis station Experion (Bio-Rad), using StdRNA chips and reagents (Bio-Rad). RNA concentrations ranged between 23 and 182 ng/μL, RQI values were >7.2.

RNAseq

Library preparation proceeded with DNase 1 digestion of total RNA, mRNA isolation by use of oligo(dT) beads, mRNA fragmentation, first-and second-strand cDNA
synthesis, end-repair, A-tailing, bar-coded adapter ligation and PCR amplification. Sequencing libraries were checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) before sequencing. Single-end (1 × 100 bp) RNA sequencing (RNASeq) data were generated using standard Illumina protocols and kits (TruSeq SBS KIT-HS v3, FC-401-3001; TruSeq SR Cluster Kit v3-cBot-HS, GD-401-3001), and all sequencing was performed using the HiSeq 2000 platforms (University of Groningen Genome Analysis Facility). Quality trimming and control

TruSeq adapters were trimmed (at a 10% error rate with cutadapt version 1.4.1, (Martin 2011)) before bases of low quality (Phred score Q < 20, 99% base call accuracy) and reads of short length (<35 bp) were removed with the FastqMcf filter in ea-utils (Aronesty 2011) (see Table S1, Supporting information for numbers of reads before and after quality trimming). Quality controls of read base content, length distribution, duplication and
over-representation were checked with fastqc http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Library NU3W2 was excluded from further analyses due to its exceptionally low number of reads (357 raw reads, compared to >10 Mio. reads in most other libraries, Table S1, Supporting information).

Mapping

We aligned reads of each library to the genome of *Z. marina* (Olsen et al. 2016) with the splice-aware RNA-seq aligner STAR (sjdbOverhang 100) (Dobin et al. 2013), guided by splice junctions from the v2.1 *Z. marina* genome annotation (gff3; GenBank Accession no.: LFYR00000000; see Table S1, Supporting information). Alignments that contained noncanonical splice junctions were filtered out. Duplicate reads were removed with the markduplicates program from the picard package (http://broadinstitute.github.io/picard/). Ambiguously mapped reads (ca. 5% in each library), defined by values >1 in the NH:i BAM file tag, were removed. NH is the number of reported alignments that contains the query in the current record. For each library, we counted the reads that mapped uniquely to annotated mRNAs (exons) with the htseq-count script of the htseq python package (Anders et al. 2015) (20 554 exons in total, see Table S2, Supporting information). Reads of low expression (library average < 5) were removed to avoid potential artefacts from sequencing errors, and reads of highly variable expression (standard deviation over all libraries > library average) were removed to reduce the effect of outlier individuals on statistical comparisons. In total, 12 948 exons remained after filtering (Table S3, Supporting information).

Annotation

Mapped sequence IDs (mRNAIDs) were associated with gene IDs, gene descriptions and Gene Ontology labels, by parsing the gff3 file of the annotated *Z. marina* genome (v2.1, nuclear and organellar, GenBank Accession no.: LFYR00000000) from the ORCAE database (Sterck et al. 2012). These gene annotations rely on inference from homology. Where gene descriptions were lacking in the gff3 file, they were transferred from functional descriptions of the top BLAST hit to *Z. marina* proteins (Table S4, Supporting information).

Population differentiation based on neutral SNPs

Neutral differentiation among the four populations was estimated from neutral SNPs (single nucleotide polymorphisms). To call SNP variants from the RNAseq data, all aligned reads were merged with samtools (Li et al. 2009) before applying gatk (McKenna et al. 2010) splitting of exon segments, reassignment of mapping qualities (SplitNCigarReads), and realignment around indels (RealignerTargetCreator and IndelRealigner). The realigned reads were demultiplexed (samtools) before calling sequence variants with gatk (HaplotypeCaller). After filtering (VariantFiltration) according to the gatk Best Practices guide for RNAseq data (http://gatkforums.broadinstitute.org/gatk/discussion/3891/calling-variants-in-rnaseq), 159 592 nuclear variants (indels and SNPs) were kept.

Variants with non-neutral divergence between the four populations were identified with the Bayesian likelihood method that is implemented in the program bayescan v2.1 (Foll & Gaggiotti 2008). The program uses differences in allele frequencies between populations to screen for non-neutral *Fₚ* outlier loci at a false discovery rate of 0.05. BayeScan approximates allele frequencies in a neutrally structured population with a multinomial Dirichlet model. Selection is introduced by decomposing *Fₚ* coefficients into a population-specific component (beta) shared by all loci, and a locus-specific component (alpha) shared by all populations using a logistic regression. This method infers posterior probabilities of each locus to be under the effect of selection by defining and comparing two alternative models (neutral vs. selection).

Nine SNPs of the 159 592 nuclear variants were identified as outlier loci with non-neutral divergence between the four populations (*q*-values <5% for the model including selection). Scaffold:locus IDs of the non-neutral variants: 1:43222, 2:501606, 2:1527432, 42:390694, 143:142100, 100:118173, 188:180877, 188:180966, 253:10530. After removing the nine outlier loci (SNPs), 20092 indels, and 170 SNPs with >2 alleles, a total of 139 321 (159592-9-20092-170) biallelic neutral SNPs (Appendix S1) were kept.

Population differentiation between the four populations was calculated from the set of 139 321 biallelic neutral SNPs as Wright’s *Fₛ*, estimated according to Weir & Cockerham (1984), and as Nei’s genetic distance (Nei 1972) with the r package ‘stamrp’ (Pemberton et al. 2013). A Neighbor-Joining (NJ) tree of Nei’s genetic distances was created and tested with 1000 bootstrap replications using the r package ‘ape’ (Paradis et al. 2004).

Multivariate clustering of gene expression

Overall transcriptomic differentiation, shaped by both neutral drift and potential selection, was characterized by clustering the samples hierarchically by the first five principle components of gene expression, averaged over technical replicates, with the principal components
analysis (PCA) and hierarchical clustering on principal components (HCPC) functions in the \texttt{R} package \texttt{FACTOMINER} (Lê et al. 2008) (setting scale.unit = FALSE not to scale the expression values to unit variance). To account for differences in sequencing depth and to assure homoscedasticity before PCA, the raw count values of mapped reads were regularized log-transformed with the function \texttt{rlog} in the \texttt{R} package \texttt{DESEQ2} (Love et al. 2014). Overall transcriptomic differentiation was characterized under control and heat stress by creating one hierarchical cluster for all control samples on the expression of all genes, and one cluster for all samples on the expression of heat-responsive genes: genes that were differentially expressed between controls and heated samples under acute stress (time points 2 and 3), or in the recovery phase (time points 5, 7 and 9) (Fig. S1, Supporting information). The first five principle components explained 76.35% of the variation in the expression of all genes in control samples and 66.96% of the variation in the expression of heat-responsive genes in all samples. Groupings of samples (NU, SU and NE (=Atlantic) vs. SE (=Mediterranean), NU and NE (=North) vs. SU and SE (=South), and controls vs. stressed and recovery samples) were tested with analysis of similarity (ANOSIM) in the \texttt{R} package \texttt{VEGAN} v2.3–1 (Oksanen et al. 2014).

\textbf{Differential expression}

We identified differences in gene expression between thermal regimes (North vs. South) by testing for differential expression in three groups of samples: (i) control, constantly kept at 19 °C at all time points; (ii) acute stress, >25 °C at time points 2 and 3; and (iii) recovery, previously warmed to >25 °C, but then allowed to recover at 19 °C and sampled at time points 5, 7 and 9 (Fig. S1, Supporting information). As the transcription profiles clearly separated the Mediterranean library (SE) from all of the Atlantic libraries (NU, SU and NE) (see hierarchical clusters in Fig. 2a,b), we also tested for differential expression between Atlantic and Mediterranean libraries. Time point was specified as additional explanatory factor to oceans (Atlantic and Mediterranean) or isotherm levels (North and South) in the differential expression tests performed with the \texttt{R}/\texttt{BIOCONDUCTOR} package \texttt{DESEQ2}. The \texttt{DESEQ2} model corrected internally for library size differences (Love et al. 2014). Significance levels of all test results were adjusted with the Benjamini and Hochberg correction (Benjamini & Hochberg 1995), using the \texttt{p.adjust} function in \texttt{R} (R Development Core Team, 2014), to control for the false discovery rate in multiple pairwise comparisons. Expression was deemed significantly different for genes with corrected \(P\)-values below 0.05.

\textbf{Acute heat-stress response and recovery}

The acute heat-stress response was determined as differential gene expression between controls and acutely stressed samples (time points 2 and 3). Recovery was determined as the differential between control expression and post-heatwave expression (time points 5, 7 and 9). Differential expression analyses were performed with the \texttt{R}/\texttt{BIOCONDUCTOR} package \texttt{DESEQ2} (Love et al. 2014) (Fig. S1, Supporting information), which internally corrected the raw count data of mapped reads for library size differences. The acute stress and the recovery responses were identified in the libraries from all populations with ‘population’ and ‘time point’ as additional explanatory factors. The acute stress response and the recovery were also identified in each population separately. While ‘time point’ was used as an additional explanatory factor to test for recovery in each population, only samples from time point 3 were used to identify the acute heat stress response for the NU and SU populations, as time point 2 samples were unavailable. Significance levels in all test results were adjusted with the Benjamini and Hochberg correction (Benjamini & Hochberg 1995), using the \texttt{p.adjust} function in \texttt{R} (R Development Core Team, 2014), to control for the false discovery rate in multiple pairwise comparisons. Expression was deemed significantly heat-responsive under acute heat stress or in the recovery phase for genes with corrected \(P\)-values below 0.05.

Enrichment tests of both gene ontologies (GO) Molecular Function (MF) and Biological Process (BP) were performed with the \texttt{R} package \texttt{TOPGO} (Alexa & Rahnenführer 2010). GO terms were obtained from the v2.1 \textit{Zostera} genome annotation from the \texttt{ORCAE} database (Sterck et al. 2012) (GenBank Accession no.: LFRY00000000). We used Fisher’s exact tests to test for enrichments in genes that were heat-responsive (significantly upregulated or downregulated in stressed vs. control samples) under acute heat stress, or in the recovery phase. To reduce redundancy in the significantly enriched GO terms (\(P\)-values < 0.05), we calculated ‘SIM REL’ scores (Schlicker et al. 2006), based on the \textit{Arabidopsis thaliana} GO-TERM database, with the \texttt{REVI戈} web server (Supek et al. 2011). The GO terms were reduced to cluster-representatives by removing terms with semantic similarities >0.5.

Previous studies identified six important ontology groups in the transcriptomic heat stress response of \textit{Z. marina}: (i) cell wall fortification (Franssen et al. 2014); (ii) protein folding and chaperone activity (Franssen et al. 2011); (iii) ribosome activity (Franssen et al. 2014); (iv) oxidation-reduction processes (Gu et al. 2012); (v) electron transport and photosynthesis (Gu et al. 2012); and (vi) osmoprotective metabolites (Street et al. 2010;
Gu et al. 2012). To estimate the representation of these ontology groups in the acute heat response and the heat response that lasted throughout the recovery phase of this study, we identified exact matches and semantic similarities (SIM REL scores < 0.5 (Schlicker et al. 2006), using REVIGO (Supek et al. 2011)) between each GO term in the enriched MFs/BPs (upregulated or downregulated under acute heat stress or recovery) and each GO term in the six targeted ontology groups (Table S5a–f, Supporting information).

**Adaptive differentiation in gene expression**

To identify signals of possible selection, we searched for genes for which the identified differential expression (North vs. South or Atlantic vs. Mediterranean, under control conditions, acute stress or recovery) could not be explained by phylogenetic distance and genetic drift alone (Fig. S1, Supporting information). This was performed using the approach of Ovaskainen et al. (2011) in the R package ‘DRIFTSEL’ 2.1.2 (Karhunen et al. 2013). Adaptive differentiation under natural selection was inferred for those genes that showed significant differential expression following phylogenetic correction under a neutral model. This was performed as follows: a matrix of population-to-population coancestry coefficients (probabilities that randomly chosen alleles for a neutral locus are identical by descent between individuals belonging to different populations) was constructed from the set of neutral biallelic SNPs with the do.all function in the ‘RAFM’ R package (Karhunen & Ovaskainen 2012), and used as a prior to estimate the posterior distributions of all parameters with a Metropolis–Hastings Markov chain Monte Carlo (MCMC) algorithm (MH function); as required to test for neutrality with the H.test function in the R package ‘DRIFTSEL’ (Karhunen et al. 2013). All Markov chains Monte Carlo converged after 3000 iterations when the Gelman-Rubin shrink factor, tracked with the R package ‘CODA’ (Plummer et al. 2006), remained close to 1. Thus, we ran a total of 6000 iterations without thinning, and discarded the first 3000 iterations as burn-in.

We used Fisher’s exact tests to test for enriched GO terms of MFs or BPs in adaptively differentiated genes (H value > 0.95) with the R package ‘TOPGO’ (Alexa & Rahnenfuhrer 2010). GO terms were based on the v2.1 Zostera genome annotation (Skerck et al. 2012) (GenBank Accession no.: LFYR00000000). To reduce redundancy in significantly enriched GO terms (P-values < 0.05 after Benjamini and Hochberg correction (Benjamini & Hochberg 1995) with the p.adjust function in R (R Core Team 2015)), we calculated ‘SIM REL’ scores (Schlicker et al. 2006), based on the Arabidopsis thaliana GO-TERM database, with the REVIGO
Coded sequence differences in temperature-adaptive genes

Twenty-one genes exhibited adaptive differentiation in gene expression exceeding neutral differentiation (H value > 0.95) between northern and southern populations, and were likely involved in the parallel adaptation of seagrass populations to warm temperatures. To test for adaptive coding sequence differences in addition to adaptive differential expression for these 21 genes, we tested for ratios of nonsynonymous to synonymous substitutions (dN/dS) exceeding 1. First, we determined the genomic consensus sequence for each population by applying population-specific SNPs to the reference genome (Olsen et al. 2016) with bcftools consensus (https://samtool.com/bcftools). Population-specific SNPs were called with GATK (McKenna et al. 2010) HaplotypeCaller and VariantFiltration from merged bam files that combined alignments of all samples from the same population. For each population, we limited the consensus sequence to the 21 target genes with bedtools getfasta (Quinlan & Hall 2010) based on the genomic features file (gff) of the Zostera genome (Olsen et al. 2016). For each target gene, codon alignments of all population sequences were obtained with pal2nal (Suyama et al. 2006) that was guided by mafft (Katoh & Standley 2013) multiple sequence alignments mafft of peptides predicted with TRANSDECODER (http://transdecoder.github.io/) based on homology to the known protein sequences.

To test for sites under positive selection in the 21 adaptively differentially expressed genes between the southern and northern populations, we performed branch-site tests by contrasting CODEML model A (relaxation, dN/dS unequal 1) to model A1 (positive selection, dN/dS > 1) of the package PAML (Yang 2007) using eTE 3 (Huerta-Cepas et al. 2016).

Results

Population differentiation based on neutral SNPs

Neutral genetic differentiation among the four populations was quantified with $F_{ST}$ values and visualized with a NJ tree. Pairwise $F_{ST}$ values ranged from 0.25 to 0.56 (all statistically significant, $P < 0.05$, Fig. 1c). The NJ tree (Fig. 1d) supports strong differentiation between European and American coasts, as well as between northern and southern populations along each coast. Notably, the Mediterranean population (SE) was the most distant from the three Atlantic populations: a common pattern associated with virtually all phylogeographic studies including seagrasses (Olsen et al. 2004).

Multivariate clustering of gene expression

Overall transcriptomic differentiation (shaped by both neutral drift and potential selection) was characterized in hierarchical clusters of gene expression with and without the impact of heat stress. Based on the expression of all mapped genes (12 948, after filtering out genes of low or highly variable expression, Table S3, Supporting information), the control samples separated into a Mediterranean (SE) and an Atlantic cluster (NU, SU and NE) (Fig. 2a). This grouping of libraries was supported by analysis of similarity ($R = 0.28, P = 0.05$). Differences in overall gene expression, thus, were in accordance with the phylogeographic divergence between the populations, represented by the Neighbor-Joining tree in Fig. 1d. In other words, a grouping of northern and southern samples in the expression of all genes was not supported by ANOSIM, $R = 0.10, P = 0.16$, Fig. 2a).

The heat-stressed samples (w, time points 2 and 3) showed a distinct expression in heat-responsive genes (4979) from the controls (c, all time points) and from the recovery samples (w, time points 5, 7 and 9) (Fig. 2b). The grouping of control and recovery samples in a separate cluster from the stressed samples was supported by ANOSIM ($R = 0.87, P = 0.001$). Atlantic and Mediterranean samples separated clearly in the control–recovery cluster (grouping supported by ANOSIM, $R = 0.25, P = 0.01$), but not in the acute stress cluster (grouping not supported by ANOSIM $R = 0.36$, $P = 0.2$), which was due to the outlier library NU3w (Fig. 2b). A grouping of northern and southern samples in the expression of heat-responsive genes was not supported by ANOSIM ($R = 0.1553, P = 0.06$, Fig. 2b).

Differential expression

Differences in gene expression between thermal regimes (North vs. South) and between oceans (Atlantic vs. Mediterranean) were identified by differential expression analysis. In each of the comparisons, the lowest number of differentially expressed genes was recorded during the acute stress phase (Table 1); the highest number of differentially expressed genes was recorded in the control samples (Table 1). The overlap of differentially expressed genes with heat-responsive genes is shown in Fig. S2a–d (Supporting information). Differentially expressed genes are listed for the Atlantic vs.
Mediterranean comparison in Table S6a–c (Supporting information), and for the North vs. South comparison in Table S6d–f (Supporting information).

**Acute heat-stress response**

The acute heat-stress response was tested as differential gene expression between controls and acutely stressed samples. NU was the only population without acute stress response. In contrast, the SU population responded at 734 genes (Table S7e, Supporting information), and the European populations responded at >1800 genes (NE: 1814, Table S7c; SE: 2004, Table S7d, Supporting information). Thus, the southern samples were not less responsive to acute heat stress than the northern samples.

A total of 4907 genes responded concordantly between all four populations to acute heat stress (Table S7a, Supporting information), based on significant differential expression between all controls and all acute stress libraries independent from population. In the acute heat stress response, 32 MFs (Table S8a, represented genes in Table S9a, Supporting information) and 46 BPs (Table S8e, represented genes in Table S9e, Supporting information) were enriched in the 1612 upregulated genes (genes with log2 fold change >0 in Table S7a, Supporting information). Dominant upregulated processes and functions, represented by >490 genes (>10% of all 4908 heat-responsive genes), included ‘cellular processes’, ‘metabolic processes’ and ‘binding’ (Table S8a,e, Supporting information). Some 38 MFs (Table S8b, represented genes in Table S9b, Supporting information) and 41 BPs (Table S8f, represented genes in Table S9f, Supporting information) were enriched in the 2395 downregulated genes (genes with log2 fold change <0 in Table S7a, Supporting information). Dominant downregulated functions and processes, represented by >490 genes (>10% of all 4908 heat-responsive genes), included ‘cellular processes’ and ‘catalytic activity’ (Tables S8b,f, Supporting information). All six BPs and MFs that were previously identified to be dominant in the heat stress response of *Z. marina* (Street et al. 2010; Franssen et al. 2011, 2014; Gu et al. 2012) (Table S5a–f, Supporting information) were also represented (semantic similarities of GO terms >0.5) in enriched heat-responsive BPs and MFs in this study: ‘Heatstress’, ‘Metabolism’, ‘Cellwall’, ‘Photosynthesis’, ‘Ribosomal’ and ‘Oxidative.reductive’ (Fig. S3a,b, Supporting information).

**Recovery**

Recovery was tested as differential gene expression between controls and recovery samples. The number of heat-responsive genes in the recovery phase was an order of magnitude lower in the southern samples (SU: 6, Table S7i; SE: 10, Table S7g, Supporting information) as compared with the northern samples (NU: 302, Table S7h; NE: 205, Table S7f, Supporting information). Given that the southern samples were not less heat-responsive than the N samples, this means that the southern samples recovered faster from heat stress.

In total, 123 genes responded concordantly between all four populations during the recovery phase (Table S7b, Supporting information). Under recovery, 12 MFs (Table S8c, represented genes in Table S9c, Supporting information) and 10 BPs (Table S8g, represented genes in Table S9c, Supporting information) were enriched in the 53 upregulated genes (genes with log2 fold change >0 in Table S7b, Supporting information), while 14 MFs (Table S8d, represented genes in Table S9d, Supporting information) and four BPs (Table S8h, represented genes in Table S9h, Supporting information) were enriched in the 70 downregulated genes (genes with log2 fold change <0 in Table S7b, Supporting information).

**Adaptive differentiation in gene expression**

We applied a phylogeographic correction to eliminate differences due to neutral processes as opposed to those due to selection. Populations were partitioned into two ways: (i) Atlantic vs. Mediterranean, and (ii) North vs. South. Atlantic and Mediterranean samples displayed the strongest adaptive signal in differential gene expression. In total, 128 of 4711 differentially expressed genes showed greater differential expression (74 under control and 106 under recovery conditions, Fig. S2a,b, Table S10a–c, Supporting information) than expected under neutral phylogenetic divergence (*H* value > 0.95), implying adaptation to the environmental covaries with a *P*-value < 0.05 (Fig. 3a). Northern and southern populations exhibited adaptive differentiation exceeding neutral differentiation (*H* value > 0.95) in 21 of 2389 differentially expressed genes (three under control and 18 under recovery conditions, Fig. S2c,d, Table S11a–c, Supporting information). None of these 21 genes showed adaptive coding sequence differences (*P*-value > 0.05 for dN/dS > 1) between northern and southern samples.
Sixteen genes exhibited adaptive differentiation in both comparisons, Atlantic vs. Mediterranean, and North vs. South (Fig. 3a,b); gene IDs based on the Z. marina genome annotation v2.1, GenBank Accession nos. LFYR00000000: Zosma5g01430, Zosma5g01440, Zosma55g00720, Zosma57g00700, Zosma68g00400, Zosma72g00300, Zosma98g00300, Zosma124g00200, Zosma21g00340, Zosma29g00070, Zosma107g00010, Zosma40g0060, Zosma425g00160, Zosma89g00800, Zosma190g00070, Zosma253g00020). None of them was adaptively differentiated under acute heat stress due to increased variation in gene expression (larger standard errors) compared to control or recovery conditions (Fig. 3b). Most of the 16 genes were lower expressed in Mediterranean and southern populations compared with Atlantic and northern populations (Fig. 3b). Thus, much of the North vs. South separation was explained by the separation between Mediterranean and Atlantic samples.

Discussion

Genes that are putatively adaptive to contrasting temperatures

Correction of differential gene expression for neutral phylogeographic differentiation enabled us to extract only the putatively adaptive portion of transcriptomic differentiation. We inferred contrasting temperatures as the major selective force when the putatively adaptive differences were correlated with temperature differences across two independent thermal clines.

The global transcriptomic differentiation (shaped by neutral phylogenetic differentiation and adaptive divergence) did not place northern and southern samples into different clusters, either under control conditions (Fig. 2a) or in response to heat stress (Fig. 2b). Nevertheless, for 21 genes (where the expression difference between northern and southern samples was greater than can be explained by phylogenetic differentiation, ca. 1% of all 2389 differentially expressed genes), adaptation by natural selection was the most parsimonious explanation. The remaining variation in these genes is most likely explained by parallel adaptation to contrasting habitat temperatures along both the American and European thermal clines. The absence of adaptive coding sequence differences \( (dN/dS < 1) \) suggests that the adaptive expression difference between northern and southern samples in these 21 genes can be ascribed to either trans-acting regulation factors or to cis-acting elements outside the coding sequences, altering gene expression regulation.

![Fig. 3 Adaptively differentiated genes.](image-url)

The Venn diagram above shows the overlap of genes that were differentially expressed (grey numbers) or adaptively differentiated (black numbers) between Atlantic (A) and Mediterranean (M) samples with those genes that were differentially expressed between northern (N) and southern (S) samples. The parallel coordinates plot below shows those 16 genes that were adaptively differentiated in both A vs. M and N vs. S comparisons, and thus, were putatively adaptive to contrasting temperatures. Coloured lines show average normalized gene expression (0–1: minimum to maximum individual expression), and shaded areas represent standard errors. Black dots indicate whether the genes were adaptively differentiated (upper dot: A vs. M, lower dot: N vs. S) under control (C), stress (S) and/or recovery (R) conditions.
Although putatively adaptive to contrasting habitat temperatures, these 21 genes may not directly affect acute-stress tolerance but instead, play a role under control or recovery conditions. This is because an increased among-sample variability in gene expression may have erased any adaptive differentiation under acute heat stress (Fig. 3b). Validation would require experimental determination of the phenotype and fitness of Zostera marina under nonstressful conditions and under recovery from heat stress (Barrett & Hoekstra 2011; Pardo-Diaz et al. 2014).

Twenty-one genes are likely a conservative representation of genes involved in adaptation to contrasting temperatures and might be extended by at least some of the genes that showed adaptive differentiation between Atlantic and Mediterranean samples. For example, 128 genes (2.8% of all 4711 differentially expressed genes) showed differential expression that could not be accounted for by neutral genetic distance in the strong transcriptionic separation between the Atlantic and Mediterranean samples (Fig. 2a,b). Additionally, two factors suggest that habitat temperature played a predominant role: (i) 76% of the genes suggesting adaptive differentiation in response to habitat temperature (16 of 21) were also adaptively differentiated between Mediterranean and Atlantic samples; and (ii) in all of these 16 genes, the directionality of differential expression agreed between southern and Mediterranean samples: under recovery, ten genes that were lower expressed in the southern samples were also lower expressed in the Mediterranean samples and six genes that were higher expressed in the southern samples were also higher expressed in the Mediterranean samples (Fig. 3b). However, the identification of genes that most likely responded to contrasting temperatures between the Mediterranean and Atlantic requires confirmation by association studies including at least one additional Mediterranean population.

The strong adaptive transcriptionic differentiation of the Mediterranean from the Atlantic samples suggests that the North vs. South differentiation of Z. marina must be stronger on the European coast than on the US Atlantic coast, and that much of the previously observed North vs. South differentiation along the European coast (NE vs. SE) (Bergmann et al. 2010; Franssen et al. 2011; Winters et al. 2011; Gu et al. 2012) might be better explained by a general Mediterranean–Atlantic (SE vs. NE, NU and SU) differentiation. The strong European North–South differentiation is likely due to high rates of genetic drift in Mediterranean populations which are small, isolated and have relatively low genetic variation (Olsen et al. 2004; Procaccini et al. 2007). Moreover, stronger North–South differentiation along the European coast is likely due to reduced gene flow (Olsen et al. 2004) favouring adaptive differentiation (Davis & Shaw 2001; Aitken et al. 2008). In contrast, on the US Atlantic coast, ongoing trans-Arctic gene flow from the E-Pacific may prevent local adaptation to warm temperatures in the South (Olsen et al. 2004). Taken together, the present study shows the strength of comparing several independent environmental clines when addressing adaptation vs. neutral differentiation in gene expression patterns.

### Local thermal adaptation in expression patterns after the heat stress

Previous common-garden experiments suggested that local thermal adaptation of European southern vs. northern populations in Z. marina was driven by faster recovery of gene expression to normal patterns after imposing a heatwave (Franssen et al. 2011). Our study confirmed that the same putatively adaptive differences in gene expression evolved in parallel along the US Atlantic coast. The finding of Franssen et al. (2011) that gene expression patterns during recovery reveal thermal adaptation better than expression patterns under acute stress was replicated on the American and European coast. Across all four populations, we found that plants recovered within one day: the gene expression of early recovery samples (taken at time point 5, one day after return to 19 °C) was indistinguishable from control samples and long-recovery samples (time points 7 and 9, 20 and 30 days after return to 19 °C, Fig. 2b). However, the extent to which populations returned to control levels of gene expression was influenced by the North–South affiliation: the southern populations expressed <20 genes differently from control levels during recovery (Table S7g,i, Supporting information), whereas the northern populations expressed >200 genes differently from control levels (Table S7f,h, Supporting information). Thus, our results show that increased stress resilience of southern seagrass samples does not only apply along the European (Franssen et al. 2011; Winters et al. 2011), but also along the North American thermal cline, suggesting reduced sensitivity to heatwaves at the species’ southern (warm) edge of distribution.

### Response to acute heat stress

Stress, as measured by the number of upregulated genes, was comparable between northern and southern populations (NU: 0; SU: 734, Table S7e; NE: 860, Table S7c; and SE: 466, Table S7d, Supporting information) and differential gene expression between all four populations was lowest during the acute stress phase (Table 1), suggesting that Z. marina relies on common
pathways to alleviate heat stress. This supports the previous work (Franssen et al. 2011) demonstrating that gene expression was not dependent on the North–South affiliation. The lack of response to acute heat stress in the American northern population (NU) is peculiar. We know that there was a heat-stress response, since it was detected during recovery conditions (Table S7h, Supporting information). However, the lack of a detectable response during acute stress might be an artefact as it is supported by a single library (all of the other acute-stress NU libraries failed, Table S1, Supporting information) that has a transcription profile differing markedly from the other acute-stress libraries (library NU3W in Fig. 2b).

Upregulation of genes involved in metabolism and cell wall synthesis most likely tempered the heat stress. In addition to the osmoprotective metabolites that were identified as an important part of the heat response in both Z. marina and Z. noltii (Gu et al. 2012), the present study found other metabolic-related genes that are known to alleviate heat stress. For example, ‘starch synthase’ (Zosma22g01480, represented in starch binding: GO:2001070, Fig. S3b, Supporting information) increased heat tolerance of wheat grains (Triticum aestivum) (Sumesh et al. 2008), and ‘glycosylation’ (GO:0070085, the post-translational attachment of carbohydrates to proteins, Fig. S3a, Supporting information) has been shown to enhance chaperone activity (Henle et al. 1998) and induced heat-shock protein synthesis in a slime mould (Murakami-Murofushi et al. 1997). Furthermore, the upregulated function ‘purine ribonucleoside binding’ (GO:0032550, Fig. S3b, Supporting information) involved 284 genes, including several stress-alleviating protein kinases (Table S9a, Supporting information).

Our results support the hypothesis of Franssen et al. (2014), that cell wall fortification may protect Z. marina from heat stress. Increased cell wall synthesis under acute stress was represented by the process ‘cellular component biogenesis’ (GO:0044085, Fig. S3a, Supporting information). Cell wall strengthening most likely continued after acute stress, as the target function ‘Cell-wall’ was represented in upregulated molecular functions during recovery (Fig. S3b, Supporting information).

Downregulation of genes related to photosynthesis and pathogen defences suggests that heat stress undermined the resistance of Z. marina to additional stress. Photosynthesis is the most heat-sensitive function in green plants (Berry & Bjorkman 1980; Weis & Berry 1988; Havaux & Tardy 1996). In our study, stress-induced photoinhibition (involving reduced carbon fixation, oxygen evolution and electron flow) was indicated by downregulated processes, such as ‘photosynthetic electron transport chain’ (GO:0009767) and ‘photosynthesis’ (GO:0015979) (Fig. S3a, Supporting information).

Pathogen defence may have been impaired by heat stress-induced downregulation of: (i) ‘cytidine deaminase’ (GO:0009972, Fig. S3a, Supporting information) and (ii) ‘cytidine deaminase activity’ (GO:0004126, Fig. S3b, Supporting information), which play important roles in the antiviral immune response through the mutagenic RNA-editing activity of cytidine deaminase (Martin et al. 2014). Rising temperatures enhance disease effects on eelgrass growth (Bull et al. 2012) and inhibit the chemical pathogen defence of eelgrass (Vergeer et al. 1995; Vergeer & Develi 1997). Rising temperatures, therefore, may indirectly increase the risk of an epidemic outbreak of the ‘wasting disease’ (Rasmussen 1977), which is caused by the protist Labyrinthula zosterae (Bockelmann et al. 2013), and triggered extensive seagrass die-offs in the 1930s and 1980s in temperate and tropical regions of the Northern Hemisphere (reviewed in Orth et al. 2006; Bishop 2013).

Resistance of Z. marina to additional anthropogenic stresses may be impaired by heat stress-induced downregulation of: (i) ‘arginine decarboxylase’ (Zosma1g02550 in ‘cellular catabolic process’ GO:0044248, Fig. S3a, Supporting information), which was also downregulated in rice (Oryza sativa) with reduced tolerance to salinity-stress (Chattopadhyay et al. 1997); (ii) ‘alpha,alpha-trehalose-phosphate synthase (UDP-forming) activity’ (GO:0003825, downregulated in the recovery phase, Fig. S3b, Supporting information), as well as associated with reduced stress tolerance in rice (O. sativa) (Li et al. 2011); and (iii) many ‘ras-related proteins’ (in ‘GTPase activity’ GO:0003924, Fig. S3b, Supporting information) that are involved in numerous aspects of cell growth control (McCormick 1995).

To conclude, the stress measured by the number of upregulated genes did not differ between southern and northern populations. The common stress response involved upregulation of genes involved in metabolism and cell wall synthesis, likely dampening the heat stress. Downregulation of genes related to photosynthesis and pathogen defence suggested that heat stress undermines the resistance of Z. marina to additional stress. Zostera marina has dominated the North Atlantic through several previous glacial–interglacial periods. Temperature alone is not the driver, but rather numerous other anthropogenic stressors press towards a tipping point.

Acknowledgements

We acknowledge Chris Weidmann and Mary Kay Fox from Waquoit Bay National Estuaries Research Reserve for permissions and assistance with sample collection from Waquoit Bay;
Fred Short from Jackson Estuarine Laboratory (UNH) for permissions and collections in Great Bay, NH, as well as links to USDA for applications; Holly Bayley, Art Mathieson and Dave Shay for field and laboratory assistance at JEL; Al Hansen for assistance with sample collection from both US sites; the Shoals Marine Laboratory, for providing housing for JLO and JAC during the US fieldwork; Philipp Schubert and Anneli Ehlers, GEOMAR/Reusch Laboratory, for assistance with sample collection from Denmark and Italy, respectively; Karsten Zecher from the Institute for Evolution and Biodiversity (University of Münster), Georg Plenge and Wilfried Niemann from the University of Münster/Bornberg-Bauer Lab, and Regina Klapper from GEOMAR/Reusch Lab for help with the mesocosm experiments and laboratory work. We are grateful to the editors Dr. Karen Chambers and Dr. Mark Ungerer as well as the two anonymous reviewers whose suggestions and comments improved the quality and clarity of the article.

This study was funded by: The Netherlands Organization for Scientific Research, Earth and Life Sciences (NWO-ALW), Project Nr. 819.01.002 to JLO and JAC (2009–2014); the Volkswagenstiftung Foundation, Evolutionary Biology Initiative for support of SF; the Alexander von Humboldt Foundation for the support of JG; the Priority Programme AQUASHIFT from the Volkmer Foundation species.

References


Vergeer LHT, Aarts TL, de Groot JD (1995) The ‘wasting disease’ and the effect of abiotic factors (light intensity, temperature, salinity) and infection with *Labyrinthula zosterae* on the phenolic content of *Zostera marina* shoots. *Aquatic Botany*, 52, 35–44.


J.L.O. led the study. T.B.H., J.A.C., S.U.F. and E.B.B. were actively involved in project planning and experimental design. J.L.O., J.A.C., N.B., S.U.F. and J.G. collected the samples. J.A.C., S.U.F., N.B. and J.G. performed the heat-stress experiments and the laboratory work. A.J. analysed the data with contributions from S.U.F. and wrote the manuscript. All co-authors read and commented on the manuscript.

Data accessibility

- Supplementary tables (Tables S1–S11) and Appendix S1 are archived at Dryad: http://dx.doi.org/10.5061/dryad.vf5fk.
- Supplementary figures (Figs S1–S3) are uploaded as Supporting Information in a single PDF file.
- RNAseq libraries will be made accessible on NCBI under BioProject number PRJNA302837.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Workflow of data analysis with colour codes representing groupings of samples/libraries.

Fig. S2 Venn diagrams showing the overlap of heat-responsive genes under different conditions.

Fig. S3 Heatmaps.
Table S1  cDNA library characteristics of all 108 cDNA libraries.

Table S2  Numbers of mapped reads.

Table S3  Regularized log-transformed expression values.

Table S4  Annotations of mapped reads.

Table S5  Targeted GO terms.

Table S6  Differential expression.

Table S7  Genes responding to heat stress.

Table S8  Enriched functions and processes under acute heat stress and in the recovery.

Table S9  Heat-responsive genes representing enriched functions and processes.

Table S10  Adaptively differentiated genes between Atlantic and Mediterranean samples.

Table S11  Adaptively differentiated genes between northern and southern samples.

Appendix S1  Biallelic neutral SNPs.