

Genome diversity in wild grasses under environmental stress

Timothy L. Fitzgerald^{a,1}, Frances M. Shapter^a, Stuart McDonald^a, Daniel L. E. Waters^a, Ian H. Chivers^b, Andre Drenth^c, Eviatar Nevo^{d,2}, and Robert J. Henry^c

^aSouthern Cross Plant Science, Plant Genomics, Southern Cross University, Lismore NSW 2480, Australia; ^bNative Seeds Pty Ltd, Cheltenham VIC 3192, Australia; ^cQueensland Alliance for Agriculture and Food Innovation, University of Queensland, Brisbane 4072, Australia; and ^dInstitute of Evolution, University of Haifa, Haifa 31905, Israel

Contributed by Eviatar Nevo, September 29, 2011 (sent for review August 15, 2011)

Patterns of diversity distribution in the *Isa* defense locus in wild-barley populations suggest adaptive selection at this locus. The extent to which environmental selection may act at additional nuclear-encoded defense loci and within the whole chloroplast genome has now been examined by analyses in two grass species. Analysis of genetic diversity in wild barley (*Hordeum spontaneum*) defense genes revealed much greater variation in biotic stress-related genes than abiotic stress-related genes. Genetic diversity at the *Isa* defense locus in wild populations of weeping ricegrass [*Microlaena stipoides* (Labill.) R. Br.], a very distant wild-rice relative, was more diverse in samples from relatively hotter and drier environments, a phenomenon that reflects observations in wild barley populations. Whole-chloroplast genome sequences of bulked weeping ricegrass individuals sourced from contrasting environments showed higher levels of diversity in the drier environment in both coding and noncoding portions of the genome. Increased genetic diversity may be important in allowing plant populations to adapt to greater environmental variation in warmer and drier climatic conditions.

adaptive variation | genomics | molecular evolution | disease resistance | abiotic stress resistance

A large volume of work has explored the relationship of genetic diversity with ecogeographic variables in plants (1). It is clear that genetic variation can result from natural selection of genotypes that are adapted to particular environments. Furthermore, different loci within any one genome may experience different selection pressures, which can result in widely varying allelic diversity among genes within the same plant species (2). Several recent studies have assessed the relationship of environmental conditions to genetic diversity in wild plant populations (3–7). Diversity measured within neutral genetic markers (5, 6) and functional loci (4, 7) has shown significant association to climatic conditions.

The *Isa* gene (8) codes for a bifunctional amylase/subtilisin inhibitor (BASI) (9), which provides defense against bacterial and fungal pathogens (10, 11) in the barley pericarp (12). BASI belongs to the proteinase inhibitor class of plant pathogenesis-related proteins (13), which inhibits proteins excreted by a range of pathogens during the infection/attack process. Diversity within the *Isa* locus in wild barley populations growing in differing environments is correlated with climatic differences, and populations from drier environments have greater diversity compared with populations growing in relatively wetter environments (14). Diversity at the *Isa* locus, especially diversity predicted to change the amino acid sequence, is likely to be significant in plant disease defense. There can be greater diversity in plant disease organisms in drier environments (15), and it may be that a diverse population of pathogens has maintained a diverse set of *Isa* alleles. Here, we studied diversity at the *Microlaena stipoides Isa* (*MIsa*) locus within an ecogeographically diverse population of weeping ricegrass [*M. stipoides* (Labill.) R. Br.], a perennial, distant wild relative of cultivated rice, to assess whether the pattern of diversity at *Isa* might be observed in another wild grass species.

A very large number of genes assist to protect plants from adverse external conditions. These defense genes include biotic

stress defense (BSD) genes, including classic R genes that interact in a gene-for-gene manner with pathogen Avr genes (16) and genes, such as *Isa*, that assist in defense against a broader spectrum of pathogens, as well as abiotic stress defense (ASD) genes that defend against stresses including drought, flooding, excess salinity, and extremes of temperature (17). Given the role of these genes at the interface of plant and environment, diversity within defense genes is of key interest to the study of molecular evolution. In this study, we have assessed the genetic diversity in additional plant defense genes [*Rpg1*, ABC1037, *Adh1*, betaine aldehyde dehydrogenase 1 (*BADH1*), and *BADH2*] with functional mechanisms that differ from the mechanisms of *Isa* in the wild barley populations studied in the work by Cronin et al. (14) to (i) compare the overall genetic diversity of these genes within identical samples from ecogeographically diverse wild populations and (ii) assess whether diversity was associated with ecogeographic variables in these populations within a broader range of plant defense genes.

Rpg1 is a classic R gene that confers resistance in barley to pathotypes of the stem rust fungus *Puccinia graminis* f. sp. *tritici* (18). ABC1037 is a closely linked but distinct, putatively functional *Rpg* family member that is hypothesized to be the progenitor of *Rpg1* because of the extremely close physical proximity and highly similar structure of the two genes (19); the highly similar structure of ABC1037 to *Rpg1* suggests that ABC1037 may perform a gene-for-gene resistance role with a yet to be identified pathogen target. *Adh1* encodes an alcohol dehydrogenase and plays a role in plant anaerobic stress defense (20–22). The two BADH isozymes encoded by *BADH1* and *BADH2* (23) are involved in proline metabolism and abiotic stress defense in plants (24), including salt stress in both barley and rice (24–26).

In addition to the nuclear genome, plants possess genomes within mitochondria and various plastids. Of these additional genomes, the chloroplast genome has been by far the most extensively studied, and the work by Green (27) provides a recent review of progress in this area. The chloroplast genome (cp-genome) is double-stranded and circular, and it is ~150 kb in most plants. A number of functional genes are encoded within the cp-genome; generally, these genes fall into three categories: photosynthesis-related genes, noncoding RNA, and *ycfs* (hypothetical proteins with poorly understood function). Multiple genome copies are present within individual chloroplasts, and many chloroplasts are present within individual plant cells (variation at both of these levels occurs during cell development and in response to environmental factors), with 1,000–10,000 cp-genome copies

Author contributions: T.L.F., F.M.S., D.L.E.W., I.H.C., E.N., and R.J.H. designed research; T.L.F., F.M.S., S.M., I.H.C., and E.N. performed research; T.L.F., F.M.S., I.H.C., E.N., and R.J.H. contributed new reagents/analytic tools; T.L.F., F.M.S., S.M., A.D., and E.N. analyzed data; and T.L.F., F.M.S., D.L.E.W., I.H.C., A.D., E.N., and R.J.H. wrote the paper.

The authors declare no conflict of interest.

¹Present address: Commonwealth Scientific and Industrial Research Organization Plant Industry, St. Lucia, Brisbane QLD, Australia.

²To whom correspondence should be addressed. E-mail: nevo@research.haifa.ac.il.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1115203108/-DCSupplemental.

present per mature photosynthetic plant cell. Compared with the nuclear genome, the cp-genome is highly conserved in terms of gene order and content, and it recombines infrequently. Genetic diversity in the chloroplast is, thus, considered extremely useful for phylogenetic analyses, facilitating comparisons over long evolutionary time scales. Advances in DNA sequencing have greatly facilitated analysis of genetic diversity at the whole-cp-genome level. As a consequence, our understanding of the dynamics of cp-genome evolution has improved dramatically in recent years (28).

Recently, a new method for cp-genome analysis has been developed (29). This method allows for cp-genome sequence analysis from total DNA and has been used to generate reference whole-cp-genome sequence for a number of Australian wild rice relatives. We have complemented the analysis of nuclear defense genes in this study by applying the method of Nock et al. (29) to the detection of chloroplast SNP diversity in weeping ricegrass samples from ecogeographically contrasting environments.

This study was undertaken to further explore genetic diversity and the association of such diversity to climatic conditions within wild grass species exposed to contrasting environmental conditions.

Results

Diversity in Wild Barley Genes. For those samples from which *Rpg1* was amplified, exon polymorphisms were detected at a rate of 2.1/100 bp, and all changes were nonsynonymous. Intron polymorphisms were detected at a rate of 1.9/100 bp (Table 1). In ABC1037, exon sequence polymorphisms were detected at 2.1/100 bp, all of which were nonsynonymous. Intron sequence polymorphisms were detected at 8.3/100 bp (Table 1). The nonsynonymous exon diversity detected in *Rpg1* and ABC1037 was reasonably evenly distributed throughout the populations, with diversity detected in five of eight populations for *Rpg1* and seven of eight populations for ABC1037 (Table S1). Only one population, Maalot, contained no functional diversity at either locus.

No polymorphisms were detected in the sequenced *Adh1* region. In *BADH1*, no exon polymorphism was detected, whereas intron polymorphisms were detected at a frequency of 0.8/100 bp (Table 1). In *BADH2*, exon sequence polymorphisms were detected at 2.2/100 bp; however, all polymorphisms were synonymous. *BADH2* intron polymorphisms were detected at 5.6/100 bp of intron sequence (Table 1). Heterozygosity was not detected in *BADH1*. Low levels of heterozygosity were detected in *BADH2*, *Rpg1*, and ABC1037. Wild barley has a high selfing rate (99%), and therefore, low levels of heterozygosity are expected. Where heterozygosity was detected, a frequency of 0.5 was attributed to both alleles at the polymorphic site.

For all loci, genotypes based on all sequence polymorphisms and nonsynonymous exon polymorphisms were determined. The number and frequency of these genotypes are displayed in Table 2. G_e values (30) based on total and nonsynonymous genotypes were calculated as a measure of total nucleotide diversity (G_eT) and encoded amino acid sequence diversity (G_eA) (Table 1). Additionally, the frequency of samples possessing the predominant genotype and the predominant encoded amino acid sequence, respectively, can be observed for all loci in Table 2.

Frequency of Synonymous vs. Nonsynonymous Polymorphism in Wild-Barley Genes. Polymorphism was detected in the exon regions of *Rpg1*, ABC1037, and *BADH2* but not *Adh1* or *BADH1*. The ratio of nonsynonymous to synonymous SNPs (*Materials and Methods*) was higher than expected in both *Rpg* family members but lower than expected in *BADH2*. For the genes assessed individually, the differences were not statistically significant, perhaps because of the relatively small number of SNPs used in the analysis. When data were combined for both *Rpg* family members, it was found that the ratio of nonsynonymous to synonymous SNPs was significantly higher than expected by chance ($P < 0.05$). Although these data were insufficient to come to any definitive conclusion, these *Rpg* genes may be under positive selection for nonsynonymous exon mutations across these populations.

Variation in Diversity Among Wild Barley Defense Loci. For each locus, varying lengths of sequence data were collected in terms of total base pairs sequenced as well as base pairs of intron and exon sequenced. Because of this variation in data, diversity within the defense loci could not be statistically compared using values calculated for G_eT and G_eA or by directly comparing the proportion of samples possessing the predominant sequence for each locus. To facilitate statistical comparison of sequence conservation within the five defense loci assessed in this study, an equation was developed to correct values for the proportion of samples possessing the predominant sequence based on the length of sequence assessed at each locus (*Materials and Methods*). Corrected, reanalyzed data for *Isa* (*Materials and Methods*) was also compared. Analyses using χ^2 contingency analysis showed that genotype and amino acid sequence conservation varied considerably among the six defense genes (Tables S2 and S3).

Diversity at *Mslsa* Locus in Weeping Ricegrass. Sequencing the *Mslsa* locus of 95 samples of weeping ricegrass from 35 different sites revealed five polymorphic sequence positions over the 350 bp that were sequenced (Table 3), and three of which were nonsynonymous. For all loci, genotypes based on all sequence polymorphism and nonsynonymous exon polymorphisms were determined (Tables 3 and 4). The frequency of each genotype is displayed in Table 2. G_e values (30) based on total and nonsynonymous genotypes were calculated as a measure of total nucleotide diversity (G_eT) and encoded amino acid sequence diversity (G_eA) (Table 1). The percentage of samples possessing the predominant *Mslsa* genotype and the predominant encoded MSISA amino acid sequence, respectively, is outlined in Tables 3 and 4.

Variation in Diversity at *Mslsa*. For the purpose of assessing whether a relationship between climatic conditions and *Mslsa* diversity existed as had been shown for *Isa* in wild barley, *Mslsa* diversity was analyzed in 35 samples, each of which originated from a different ecogeographic location (Fig. S1). Samples were grouped based on climatic conditions as colder, wetter (CW; mean annual temperature $< 19^\circ\text{C}$ and mean annual rainfall $> 1,000$ mm) and hotter, drier (HD; mean annual temperature $> 19^\circ\text{C}$ and mean annual rainfall $< 1,000$ mm). Samples from 89.5% of CW sites possessed a single predominant genotype,

Table 1. Polymorphism and diversity detected in defense loci

Gene	Sample number	Intron polymorphisms/100 bp	Synonymous exon polymorphisms/100 bp	Nonsynonymous exon polymorphisms/100 bp	G_eT	G_eA
<i>Mslsa</i>	95	N/A	0.6 (2)	0.9 (3)	0.54	0.32
<i>Rpg1</i>	44	1.9	0	2.1 (8)	0.73	0.58
ABC1037	91	8.3	0	2.1 (6)	0.76	0.44
<i>Adh1</i>	94	0	0	0	0	0
<i>BADH1</i>	91	0.8	0	0	0.14	0
<i>BADH2</i>	87	5.6	2.2 (3)	0	0.79	0

Genotypic diversity was calculated using the diversity index by Nei (30) for all loci based on all polymorphism (G_eT) and nonsynonymous exon polymorphism (G_eA). For exon mutations, the absolute number of polymorphisms is given in parentheses, and therefore, the ratio of synonymous to nonsynonymous polymorphisms can be observed.

Table 2. Genotypes (GEN) and frequencies for defense loci based on all polymorphism and nonsynonymous (NS) exon polymorphism

Gene	All polymorphism		NS exon polymorphism	
	GEN	Frequency	GEN	Frequency
<i>Mslsa</i>	1	0.64	1	0.80
<i>Mslsa</i>	2	0.02	2	0.013
<i>Mslsa</i>	3	0.015	3	0.003
<i>Mslsa</i>	4	0.003	4	0.003
<i>Mslsa</i>	5	0.003	5	0.18
<i>Mslsa</i>	6	0.003		
<i>Mslsa</i>	7	0.17		
<i>Mslsa</i>	8	0.14		
<i>Mslsa</i>	9	0.008		
<i>Rpg1</i>	1	0.48	1	0.61
<i>Rpg1</i>	2	0.09	2	0.16
<i>Rpg1</i>	3	0.14	3	0.11
<i>Rpg1</i>	4	0.11	4	0.05
<i>Rpg1</i>	5	0.05	5	0.02
<i>Rpg1</i>	6	0.05	6	0.02
<i>Rpg1</i>	7	0.02	7	0.02
<i>Rpg1</i>	8	0.02		
<i>Rpg1</i>	9	0.02		
<i>Rpg1</i>	10	0.02		
ABC1037	1	0.42	1	0.72
ABC1037	2	0.13	2	0.15
ABC1037	3	0.15	3	0.13
ABC1037	4	0.10		
ABC1037	5	0.10		
ABC1037	6	0.03		
ABC1037	7	0.04		
ABC1037	8	0.01		
ABC1037	9	0.01		
<i>Adh1</i>	1	1.00	1	1.00
<i>BADH1</i>	1	0.92	1	1.00
<i>BADH1</i>	2	0.06		
<i>BADH1</i>	3	0.01		
<i>BADH1</i>	4	0.01		
<i>BADH2</i>	1	0.32	1	1.00
<i>BADH2</i>	2	0.28		
<i>BADH2</i>	3	0.14		
<i>BADH2</i>	4	0.1		
<i>BADH2</i>	5	0.03		
<i>BADH2</i>	6	0.06		
<i>BADH2</i>	7	0.01		
<i>BADH2</i>	8	0.02		
<i>BADH2</i>	9	0.02		
<i>BADH2</i>	10	0.01		
<i>BADH2</i>	11	0.01		

whereas only 31.5% of the samples from HD sites possessed the predominant genotype (Table 5). Furthermore, 97.5% of samples from CW sites possessed a single predominant encoded amino acid sequence, whereas only 81.7% of samples from HD sites possessed the predominant encoded amino acid sequence (Table 5). The differences in genotypic and encoded amino acid sequence conservation between groups were both highly significant ($P < 0.001$ using a Fisher exact two-sided test). Furthermore, eight individual genotypes and five individual encoded amino acid sequences were identified among the samples from HD sites, whereas only five individual genotypes and two individual encoded amino acid sequences were identified among the samples from CW sites. Diversity as measured by both G_eT and G_eA was also substantially higher within HD than CW samples (Table 5).

Chloroplast Diversity in an Individual Weeping Ricegrass Sample.

Next generation sequence data from a single weeping ricegrass

plant (29) was analyzed to assess diversity among the chloroplast genome copies within a single individual. The cp-genome SNP density within the individual chosen for analysis was 0.063 SNP/100 bp. This finding was determined from a reference assembly that aligned 486,494 reads to the chloroplast genome sequence, resulting in average coverage of 121 \times .

Variation in Chloroplast Diversity Between Weeping Ricegrass Populations.

Next generation sequencing of the two equimolar pools of 11 individual weeping ricegrass samples representing the two environmentally diverse sites (site A is hot, dry coastal and site B is cooler, wetter semi-alpine) resulted in 67,765,033 reads for site A and 65,440,317 reads for site B. Alignment of these reads independently with a weeping ricegrass chloroplast genome reference sequence (GenBank accession no. GU592211) with a minimum coverage of 88 \times for any reference position in both pools required resulted in 75,096 bp of the chloroplast reference sequence being evaluated with an average coverage of 1,232 \times at site A and 854 \times at site B.

To ensure robustness of the SNP analysis in addition to the quality stringencies applied to the alignment and SNP detection (*Materials and Methods*), an additional requirement of a minimum frequency of 5% for any SNP was determined to be optimal to exclude erroneous identification of polymorphisms generated by the alignment of homologous nuclear genomic sequence to the chloroplast reference. The SNP density established for the pooled samples was 1.01 SNPs/100 bp for site A and 0.76 SNPs/100 bp for site B (Table 6). A Fisher exact test of the proportion of SNP at site A compared with site B indicates that there is a significantly higher proportion of SNP at site A ($P < 0.00001$). Although the two collection sites were geographically isolated and had very different environmental conditions, there were 459 SNPs common to both sites, with 298 and 111 SNPs specific to sites A and B, respectively. Of the SNPs identified, nonsynonymous SNPs occurred at a rate of 0.197/100 bp at site A and 0.165/100 bp at site B (Table 6).

Discussion

***Isa* and *Mslsa* Diversity.** *Isa* diversity in the wild barley population assessed in this study was previously determined to be relatively high and strongly associated with water availability (14). Although *Isa* is involved in disease defense, it is unlikely to have coevolved in a gene-for-gene manner. Instead, BASI is a bifunctional inhibitor of α -amylases and serine proteases of the subtilisin family (31) and seems to inhibit proteins excreted by a range of pathogens during the infection/attack process. Greater diversity at this locus within samples from relatively dry environments may reflect a greater diversity of microbial populations (14) against which *Isa* defends. It has been hypothesized that environments where water is not limiting may support numerically larger but less genetically diverse microbial populations (14, 15). Arid environments typically experience extreme fluctuations in water availability that can lead to diverse soil microclimates (32); such diverse microclimates represent ecological niches that encourage microbial species diversity (33). Furthermore, the evolutionary rate of plant pathogens has, in some circumstances, been shown to be increased in response to elevated temperature (34) and increased water stress (35).

Weeping ricegrass [*M. stipoides* (Labill.) R. Br.] is a distant wild relative of rice that is widely distributed throughout eastern Australia. *Mslsa* was more diverse in samples from relatively drier environments, suggesting the phenomenon operating in wild barley at this locus may be found more generally throughout the grasses. It is intriguing to speculate whether *Isa* and *Mslsa* could provide defense against similar plant pathogens within these geographically isolated, distinct wild grass species. Pathogens of *Microalaena* are poorly studied, although there is some overlap between diseases of *Microalaena* and diseases of cereal crops, including barley [e.g., ergot (*Claviceps purpurea* and *C. phalaridis*)] (36).

Table 3. Observed genotype variation within *Mslsa* across the weeping ricegrass population

Genotype	Base position in consensus sequence from all individuals					All samples (n = 95)	Samples from unique sites (n = 35)
	30	99	126	129	199		
1	C	C	G	T	C	60.75	24.75
2	T	C	G	T	C	2	2
3	C	A	G	G	C	1.25	0.25
4	C	A	G	C	C	0.25	0.25
5	C	C	G	T	C	0.25	0.25
6	C	C	G	C	C	0.25	0.25
7	C	C	A	T	C	16.25	2.25
8	C	C	G	T	T	13.25	4.75
9	C	C	A	T	T	0.75	0.25
Putative amino acid change	Syn	Pro to Glu*	Arg to His	Leu to Arg*	Syn	63.9% predominant sequence	70.7% predominant sequence

Note that heterozygous individuals have been counted as a representative percentage of the individual for a given haplotype. Where two heterozygous positions were present in the same haplotype, all four resultant putative haplotypes were assigned 0.25 of an individual. Note that the column of samples from unique sites is a subset of the column of all samples.

*Change of amino acid structure is likely to affect protein formation.

Wild Barley Defense Genes. To further explore diversity in plant defense loci and its association with environment, we analyzed DNA sequence within functionally significant regions of five additional defense genes (*Rpg1*, *ABC1037*, *Adh1*, *BADH1*, and *BADH2*) in the wild barley populations assessed in the work by Cronin et al. (14). To statistically compare the degree of diversity present within the defense genes, we developed a method that accommodates variation in length of sequence and number of samples assessed between loci. Furthermore, we reanalyzed *Isa* diversity detected in the work by Cronin et al. (14) within a functionally significant gene region (*Materials and Methods*) in the samples assessed here and compared this diversity to diversity present in the other defense genes. Using this method, we observed highly significant differences in frequencies of the most abundant genotypes (both in terms of total DNA polymorphism and encoded amino acid sequence) among the defense genes (Tables S2 and S3), which strongly suggest that the rate at which molecular evolution is occurring varies considerably among these genes. The bias to nonsynonymous SNPs observed in *Rpg1* and *ABC1037* suggests that these genes may be under selection for functional diversification in this population. However, no significant association between diversity within any of the additional defense genes assessed in this study and any of the ecogeographic variables was observed here, indicating that such association is locus-specific.

Plant defense genes can be assigned to the broad categories of BSD (i.e., disease resistance) and ASD. Multiple allelic sequences encoding unique amino acid sequences were identified for each of the BSD genes (*Isa*, *Rpg1*, and *ABC1037*) in these samples (Table 2). In stark contrast, only one encoded amino

acid sequence was identified for each of the ASD genes (*Adh1*, *BADH1*, and *BADH2*) (Table 2). It is tempting to speculate that this observation may be indicative of a broad trend in ASD vs. BSD gene diversity, because it is intuitively appealing to assume that different selective pressures would generally act on BSD and ASD genes. It seems likely that many alleles of a biotic defense gene would exist throughout a diverse plant population in response to genetic diversity within the disease-causing agents against which it provides defense. In contrast, even in an ecogeographically diverse wild plant population, abiotic stress varies only in degree, and therefore, it seems possible that a single, highly effective genotype could dominate, which has been observed in the ASD genes assessed in this study.

Weeping Ricegrass Chloroplast Diversity. Chloroplast sequence diversity has been widely used as measure of plant genetic diversity and phylogeny (27); in the past, chloroplast diversity has been assessed using a single or small number of genes. With recent advances in DNA sequencing, the analysis of chloroplast sequence diversity at the whole-genome level has become feasible (27). Recently, the work by Nock et al. (29) showed a method for chloroplast genome sequencing from total DNA using the Illumina GAII system. We recognized that this method was applicable to SNP diversity analysis within pooled DNA samples and have optimized and applied this strategy here. For any method of analysis of sequence diversity, the determination of an appropriate frequency at which to consider a polymorphism real (as distinguished from sequencing and/or bioinformatics error or other artifact) is critical. Using this method, potentially significant sources of error include biases in sample preparation and error rates of the

Table 4. Observed encoded amino acid sequence variation within *Mslsa* across the weeping ricegrass population

Genotype	Base position in consensus sequence from all individuals			All samples (n = 95)	Samples from unique sites (n = 35)
	99	126	199		
1	C	G	T	76.25	31.75
2	A	G	T	1.25	0.25
3	A	G	G	0.25	0.25
4	C	G	G	0.25	0.25
5	C	A	T	17	2.5
Putative amino acid change	Pro to Glu*	Arg to His	Leu to Arg*	80.3% predominant sequence	90.7% predominant sequence

Note that heterozygous individuals have been counted as a representative percentage of the individual for a given haplotype. Where two heterozygous positions were present in the same haplotype, all four resultant putative haplotypes were assigned 0.25 of an individual. Note that the column of samples from unique sites is a subset of the column of all samples.

*Change of amino acid structure is likely to affect protein formation.

Table 5. Distribution of sequence diversity in *Mslsa* between uniquely sampled hot, dry (HD), and cold wet (CW) sites

Site climate	PG (%)	PAA (%)	G_eT	G_eA
CW	89.5	97.5	0.16	0.05
HD	31.5	81.7	0.70	0.31

PAA, predominant amino acid; PG, predominant genotype.

sequencing instrument as well as the detection of sequence from the nuclear genome. In regard to the latter source, assembly against a chloroplast reference of massively parallel sequenced genomic DNA will also align any homologous nuclear genome sequence. It is well documented that transfer of chloroplast sequence to the nucleus is evolutionarily ongoing and that numerous nuclear plastid sequences (NUPTs) occur within the nuclear plant genome (37). Generally, NUPTs are highly conserved relative to their corresponding chloroplast sequences, and recent studies have quantified the majority as being between 97% and 100% homologous with the chloroplast reference sequence (38, 39). Any polymorphism occurring between NUPTs and the respective chloroplast sequence will potentially present as an SNP using this method. However, because of the thousands of chloroplast genome copies present per cell compared with the single nuclear genome (27), the quantity of true chloroplast gene sequences will be several orders of magnitude higher than any NUPT, and thus, frequency of any individual erroneous SNP between genomes will be low. Taking these issues into consideration, we applied a cutoff of 5% frequency for SNP to be implemented into the analysis of chloroplast genome diversity.

SNP density within the chloroplast genome was significantly different within pooled weeping ricegrass samples from contrasting ecogeographic environments. Pooled chloroplast genome sequences from samples at the relatively hot, dry coastal site featured a significantly higher frequency of SNP than sequences from samples at the cooler, wetter semi-alpine site. A relatively high proportion of the chloroplast genome encodes functional genes, and additional analysis showed that ~20% of the detected SNPs were nonsynonymous and therefore, potentially functionally significant. Nonsynonymous SNPs were significantly more frequent at the hotter, drier site as well. This pattern of environment-related diversity at the level of the whole-chloroplast genome parallels the nuclear locus-specific trends observed in *Isa* from wild barley and *Mslsa* from these weeping ricegrass populations. Higher diversity within the cp-genome [measured using simple sequence repeat (SSR) loci] from wild barley samples exposed to relatively hotter and drier climatic conditions has also previously been observed (40).

Conclusions

This study reinforces the fundamental biological concept that climatic conditions can be associated with genetic diversity within wild plant populations (1), but it also highlights that genetic diversity and its association with environmental conditions can be highly locus-specific. *Mslsa* diversity was significantly higher in

Table 6. SNP distribution in the chloroplast genome identified by pooled next generation resequencing from two isolated weeping ricegrass populations

	Individual sample	Site A	Site B
Total SNPs identified	57	757	570
SNPs/100 bp	0.063	1.008	0.759
Common SNPs	30	459	459
Unique SNPs	27	298	111
Noncoding SNPs	—	352	272
Coding synonymous SNPs	—	247	174
Nonsynonymous SNPs	—	148	124
Nonsynonymous SNPs/100 bp	—	0.197	0.165

Site A (hot, dry coastal locality) and site B (wet, cool semi-alpine locality).

weeping ricegrass samples from hotter, drier sites, and this pattern reflects that observed in *Isa* within wild barley samples (14). However, no significant association of diversity to climatic conditions was observed within five additional defense loci from wild barley, although highly variable levels of diversity were detected within the genes. *Adh1*, *BADH1*, and *BADH2* are involved in plant defense against abiotic stress. Like *Isa*, *Rpg* family members function in biotic defense; however, they do so in a distinct manner. Although *Isa* seems to provide broad spectrum defense by inhibition of pathogen proteinases, *Rpg1* and (putatively) ABC1037 are R genes encoding proteins that interact with specific pathogen Avr proteins. These key functional differences may explain the observed differences in climatic association of diversity among these genes.

In this study, whole-chloroplast genome diversity has been assessed for associations with climatic conditions, and total SNP diversity within the chloroplast genome was also significantly higher in weeping ricegrass samples exposed to hotter, drier climatic conditions. Given the widespread use of chloroplast diversity as a measure of plant genetic diversity (27), these results could be extrapolated to suggest that weeping ricegrass samples exposed to hotter, drier climatic conditions are generally more genetically diverse. However, the locus-specific nature of genetic diversity and its climatic association, which could be inferred from studies in the past and have been elucidated here, prescribe caution for such extrapolation.

With access to increasingly powerful sequencing technology, the definition of total sequence diversity—within the nuclear genome and the genomes of all plant organelles as well as within the transcriptome—from multiple samples in diverse plant populations may soon be within reach. A very large proportion of plant DNA sequence is repetitive and/or noncoding, and recent studies suggest that transposons/retrotransposons can display ecogeographic association (41). It is, therefore, extremely interesting to speculate in regard to how widespread associations of molecular diversity to environmental conditions are within wild plant populations (42). New tools for molecular analysis will certainly be critical to aid in future analyses, and here, we have presented two such tools. The first tool is an adaptation of the method of Nock et al. (29) for detecting chloroplast SNP from bulked whole DNA samples using massively parallel sequencing technology. The second tool is an equation that facilitates the statistical comparison of DNA/amino acid sequence conservation among loci that vary in length.

This study provides additional insight into the molecular evolution of wild crop relatives exposed to diverse climatic conditions. Such insight may assist in the development of crop plants that are better adapted to a future climate that is hotter and drier.

Materials and Methods

Additional details are provided in *SI Materials and Methods*.

Plant Materials. Wild barley (*Hordeum spontaneum*) samples from eight genetically distinct populations of wild barley sourced from discrete localities throughout Israel were analyzed. Each locality is associated with specific ecogeographical variables (Table S4). From the 18–30 samples from each of eight ecogeographically diverse locations within Israel that were used in a study of *Isa* diversity (14), 11 or 12 samples for each population were arbitrarily selected for analysis.

Weeping ricegrass (*M. stipoides*) samples were collected on a 238.6-km east to west transect running from Studley Park Boathouse, inner Melbourne, Australia, to Paynesville on the Gippsland coast of Victoria, Australia (Fig. S2 and Table S5). The transect was within latitude 37.77°S and 38.02°S (a range of 0.24°), and the altitude ranged from just above sea level at the eastern and western extremes of the transect to the permanent snowline of Mount Baw Baw (939 m) (Fig. S1). The populations collected were not continuous. There were 37 sample sites that varied from ~1 to ~35 km apart. Leaf samples were collected from two plants at each location. Global positioning system location and elevation data were recorded with a description of the site and soil characteristics. Soil was analyzed for pH using a Manutec soil pH test kit. Leaf samples were collected from an additional 10 plants at six of the sampling sites (at altitudes of 3, 58, 317, 497, 882, and 939 m).

Accession numbers for samples used in this study are presented in Table S6.

PCR and Sequencing. Primers for both PCR amplification and terminator sequencing reactions (Table S7) of *BADH1*, *BADH2*, *Adh1*, ABC1037, and *Rpg1* were designed based on the available sequence from both cultivated and wild barley using the program Primer Premier Version 5.0 (Premier Biosoft International). Primers designed to amplify and sequence *Mslsa* were based on conserved regions of an alignment of *Oryza sativa* and *Hordeum vulgare* *Isa* gene sequence (Table S7). *SI Materials and Methods* has additional details.

Reanalysis of Diversity at the *Isa* Locus. The data obtained in the work by Cronin et al. (14) for the *Isa* locus in this population were reanalyzed for this study. Samples (11 or 12) from each of the populations were arbitrarily selected (without any consideration of diversity identified at this locus in any of the individual samples) for this analysis. A 200-bp (average length of exon sequence assessed in the other defense loci) section from the region of the *Isa* cDNA sequence between 150 and 500 bp, which encodes a number of key amino acids acting at the interface between BASI and α -amylase 2 (AMY2) (43, 44), was selected, and polymorphisms were analyzed (Fig. S3).

Correction of Proportion of Samples Possessing Predominant Sequence. An equation was developed to correct values for the percentage of samples possessing the predominant sequence (i.e., genotype or encoded amino acid sequence) based on the length of sequence assessed to allow for comparison of sequence conservation among defense loci (*SI Materials and Methods* has detailed explanation of these calculations) (Eq. 1):

$$A_y = A_x(A_x/B)^{(y/x-1)}, \quad [1]$$

where total sample number = B , length of sequence assessed in base pairs = x , samples possessing predominant sequence from x length of sequence =

- Nevo E (2001) Evolution of genome-phenome diversity under environmental stress. *Proc Natl Acad Sci USA* 98:6233–6240.
- Jing R, et al. (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). *Genetics* 177:2263–2275.
- Badri M, Zitoun A, Illahi H, Huguet T, Aouani ME (2008) Morphological and microsatellite diversity associated with ecological factors in natural populations of *Medicago laciniata* Mill. (Fabaceae). *J Genet* 87:241–255.
- Wang J-R, et al. (2008) Molecular evolution of dimeric alpha-amylase inhibitor genes in wild emmer wheat and its ecological association. *BMC Evol Biol* 8:91.
- Dong P, et al. (2009) EST-SSR diversity correlated with ecological and genetic factors of wild emmer wheat in Israel. *Hereditas* 146:1–10.
- Peleg ZVI, et al. (2008) Allelic diversity associated with aridity gradient in wild emmer wheat populations. *Plant Cell Environ* 31:39–49.
- Yang Z, Zhang TAO, Bolshoy A, Beharav A, Nevo E (2009) Adaptive microclimatic structural and expressional dehydrin 1 evolution in wild barley, *Hordeum spontaneum*, at 'Evolution Canyon', Mount Carmel, Israel. *Mol Ecol* 18:2063–2075.
- Hejgaard J, Bjørn SE, Nielsen G (1984) Localization to chromosomes of structural genes for the major protease inhibitors of barley grains. *Theor Appl Genet* 68:127–130.
- Leah R, Mundy J (1989) The bifunctional α -amylase/subtilisin inhibitor of barley: Nucleotide sequence and patterns of seed-specific expression. *Plant Mol Biol* 12:673–682.
- Mundy J, Svendsen IB, Hejgaard J (1983) Barley α -amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Res Commun* 48:81–90.
- Sancho AI, et al. (2003) Cross-inhibitory activity of cereal protein inhibitors against α -amylases and xylanases. *Biochim Biophys Acta* 1650:136–144.
- Furtado A, Henry R, Scott K, Meech S (2003) The promoter of the *Asi* gene directs expression in the maternal tissues of the seed in transgenic barley. *Plant Mol Biol* 52:787–799.
- Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC (2008) Plant pathogenesis-related (PR) proteins: A focus on PR peptides. *Plant Physiol Biochem* 46: 941–950.
- Cronin JK, Bundock PC, Henry RJ, Nevo E (2007) Adaptive climatic molecular evolution in wild barley at the *Isa* defense locus. *Proc Natl Acad Sci USA* 104:2773–2778.
- Kuang H, van Eck HJ, Sicard D, Michelmore R, Nevo E (2008) Evolution and genetic population structure of prickly lettuce (*Lactuca serriola*) and its RGC2 resistance gene cluster. *Genetics* 178:1547–1558.
- Bent AF, Mackey D (2007) Elicitors, effectors, and R genes: The new paradigm and a lifetime supply of questions. *Annu Rev Phytopathol* 45:399–436.
- Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: Achievements and limitations. *Curr Opin Biotechnol* 16:123–132.
- Brueggeman R, et al. (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. *Proc Natl Acad Sci USA* 99:9328–9333.
- Brueggeman R, Drader T, Kleinhofs A (2006) The barley serine/threonine kinase gene *Rpg1* providing resistance to stem rust belongs to a gene family with five other members encoding kinase domains. *Theor Appl Genet* 113:1147–1158.
- Lemke-Keyes CA, Sachs MM (1989) Anaerobic tolerant null: A mutant that allows *Adh1* nulls to survive anaerobic treatment. *J Hered* 80:316–319.
- Matsumura H, Takano T, Takeda G, Uchimiya H (1998) *Adh1* is transcriptionally active but its translational product is reduced in a rad mutant of rice (*Oryza sativa* L.), which is vulnerable to submergence stress. *Theor Appl Genet* 97:1197–1203.
- Licausi F, Perata P (2009) Low oxygen signaling and tolerance in plants. *Advances in Botanical Research*, eds Jean-Claude K, Michel D (Academic, London), Vol 50, pp 139–198.
- Nakamura T, et al. (2001) An isozyme of betaine aldehyde dehydrogenase in barley. *Plant Cell Physiol* 42:1088–1092.
- Fitzgerald TL, Waters DLE, Henry RJ (2009) Betaine aldehyde dehydrogenase in plants. *Plant Biol (Stuttg)* 11:119–130.
- Fitzgerald TL, Waters DLE, Henry RJ (2008) The effect of salt on betaine aldehyde dehydrogenase transcript levels and 2-acetyl-1-pyrroline concentration in fragrant and non-fragrant rice (*Oryza sativa*). *Plant Sci* 175:539–546.
- Fitzgerald TL, Waters DLE, Brooks LO, Henry RJ (2010) Fragrance in rice (*Oryza sativa*) is associated with reduced yield under salt treatment. *Environ Exp Bot* 68:292–300.
- Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. *Plant J* 66:34–44.
- Clarke JL, Daniell H, Nugent JM (2011) Chloroplast biotechnology, genomics and evolution: Current status, challenges and future directions. *Plant Mol Biol* 76:207–209.
- Nock CJ, et al. (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol J* 9:328–333.
- Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321–3323.
- Bellincampi D, et al. (2004) Potential physiological role of plant glycosidase inhibitors. *Biochim Biophys Acta* 1696:265–274.
- Iogna PA, Bucci SJ, Scholz FG, Goldstein G (2011) Water relations and hydraulic architecture of two Patagonian steppe shrubs: Effect of slope orientation and microclimate. *J Arid Environ* 75:763–772.
- Achtman M, Wagner M (2008) Microbial diversity and the genetic nature of microbial species. *Nat Rev Microbiol* 6:431–440.
- Pangga IB, Hanan J, Chakraborty S (2011) Pathogen dynamics in a crop canopy and their evolution under changing climate. *Plant Pathol* 60:70–81.
- Newton A, Johnson S, Gregory P (2011) Implications of climate change for diseases, crop yields and food security. *Euphytica* 179:3–18.
- Walker J (2004) *Claviceps phalaridis* in Australia: Biology, pathology and taxonomy with a description of the new genus *Cepsiclava* (Hypocreales, Clavicipitaceae). *Austr Plant Path* 33:211–239.
- Bock R, Timmis JN (2008) Reconstructing evolution: Gene transfer from plastids to the nucleus. *Bioessays* 30:556–566.
- Hasegawa M, Zhong B, Zhong Y (2009) Adaptive evolution of chloroplast genomes in ancestral grasses. *Plant Signal Behav* 4:623–624.
- Kleine T, Maier UG, Leister D (2009) DNA transfer from organelles to the nucleus: The idiosyncratic genetics of endosymbiosis. *Annu Rev Plant Biol* 60:115–138.
- Nevo E, et al. (2005) Genomic microsatellite adaptive divergence of wild barley by microclimatic stress in 'Evolution Canyon', Israel. *Biol J Linn Soc Lond* 84:205–224.
- Nevo E (2009) Evolution in action across life at "Evolution Canyons", Israel. *Trends Evol Biol* 1:e3.
- Nevo E, Beiles A (2011) Genetic variation in nature. *Scholarpedia* J 6:8821.
- Bonsager BC, et al. (2005) Mutational analysis of target enzyme recognition of the β -trefoil fold barley α -amylase/subtilisin inhibitor. *J Biol Chem* 280:14855–14864.
- Michielsens PO, et al. (2008) Structural and mutational analyses of the interaction between the barley α -amylase/subtilisin inhibitor and the subtilisin savinase reveal a novel mode of inhibition. *J Mol Biol* 380:681–690.