77

BARD

FINAL REPORT
PROJECT NO. US-2547-95

**Genetic Diversity of Resistance Gene Clusters
in Wild Populations of *Lactuca***

R. Michelmore, E. Nevo, A. Korol, T. Fahima

2000

BARD Final Scientific Report

(Cover Page)

Date of Submission of the report: Feb 15, 2000

BARD Project Number: US-2547-95

Evaluating Panel:

Project Title: Genetic diversity of resistance gene clusters in wild populations of *Lactuca*.

Address & e-mail of Investigators and Institutions:

Investigators

Institutions

Principal Investigator:

Richard Micheltmore (rwmicheltmore@ucdavis.edu)

University of California, Davis

Cooperating Investigators:

Ebi Nevo (nevo@research.haifa.ac.il)

University of Haifa, Israel

Abraham Korol (rab1307@uvm.haifa.ac.il)

University of Haifa, Israel

Tzion Fahima

University of Haifa, Israel

Continuation of (Related to) Previous BARD Project:

☐ Yes

☒ No

Number:

Keywords *not* appearing in the title and in order of importance. Avoid abbreviations.

Gene conversion, unequal crossing-over, molecular marker

Abbreviations used in the report, in alphabetical order:

AFLP: Amplified Fragment Length Polymorphism

Budget: IS: \$150,000

US: \$150,000

Total: \$300,000

Signature
Principal Investigator

MAR 01 2000

Signature
Research Authority, Principal Institution
Fay Yee
Contract & Grant Analyst

Publication Summary (numbers)

	Joint IS/US authorship	US Authors only	Israeli Authors only	Total
Refereed (published, in press, accepted)	1			1
Submitted, in review, in preparation	3			3
Invited review papers		1		1
Book chapters				0
Books				0
Master theses				0
Ph.D. theses				0
Abstracts	2			2
Not refereed (proceedings, reports, etc.)				0

Patent Summary (numbers)

	Israeli inventor only	US inventor only	Joint IS/US inventors	Total
Submitted				0
Issued (allowed)				0
Licensed				0

Cooperation Summary (numbers)

	From US to Israel	From Israel to US	Together, elsewhere	Total
Visits/Meetings	0	2	0	2
Sabbaticals				0
Postdoctorates	3	2		5

- **Cooperation**, briefly explain whether synergistic, complementary or supportive.

Complementary and supportive depending on the aspect of the project. Complementary plant populations were sampled in California and Israel and the two groups were mutually supportive in terms each contributing different strengths: marker technology from the US and population genetic expertise from Israel.

Abstract

Genetic resistance is often the least expensive, most effective, and ecologically-sound method of disease control. It is becoming apparent that plant genomes contain large numbers of disease resistance genes. However, the numbers of different resistance specificities within a genepool and the genetic mechanisms generating diversity are poorly understood. Our objectives were to characterize diversity in clusters of resistance genes in wild progenitors of cultivated lettuce in Israel and California in comparison to diversity within cultivated lettuce, and to determine the extent of gene flow, recombination, and genetic instability in generating variation within clusters of resistance genes.

Genetic diversity of resistance genes was analyzed in wild and cultivated germplasm using molecular markers derived from lettuce resistance gene sequences of the NBS-LRR type that mapped to the major cluster of resistance genes in lettuce (Sicard *et al.* 1999). Three molecular markers, one microsatellite marker and two SCAR markers that amplified LRR-encoding regions, were developed from sequences of resistance gene homologs at the *Dm3* cluster (*RGC2s*) in lettuce. Variation for these markers was assessed in germplasm including 74 genotypes of cultivated lettuce, *L. sativa* and 71 accessions of the three wild *Lactuca* spp., *L. serriola*, *L. saligna* and *L. virosa* that represent the major species in the sexually accessible genepool for lettuce. Diversity was also studied within and between natural populations of *L. serriola* from Israel and California. Large numbers of haplotypes were detected indicating the presence of numerous resistance genes in wild species.

We documented a variety of genetic events occurring at clusters of resistance genes for the second objective (Sicard *et al.*, 1999; Woo *et al.*, in prep; Kuang *et al.*, in prep^b). The diversity of resistance genes in haplotypes provided evidence for gene duplication and unequal crossing over during the evolution of this cluster of resistance genes. Comparison of nine resistance genes in cv. *Diana* identified 22 gene conversion and five intergenic recombinations. We cloned and sequenced a 700 bp region from the middle of *RGC2* genes from six genotypes, two each from *L. sativa*, *L. serriola*, and *L. saligna*. We have identified over 60 unique *RGC2* sequences. Phylogenetic analysis surprisingly demonstrated much greater similarity between than within genotypes. This led to the realization that resistance genes are evolving much slower than had previously been assumed and to a new model as to how resistance genes are evolving (Michelmore and Meyers, 1998).

The genetic structure of *L. serriola* was studied using 319 AFLP markers (Kuang *et al.*, in prep^a). Forty-one populations from Turkey, Armenia, Israel, and California as well as seven European countries were examined. AFLP marker data showed that the Turkish and Armenian populations were the most polymorphic populations and the European populations were the least. The Davis, CA population, a recent post-Columbian colonization, showed medium genetic diversity and was genetically close to the Turkish populations. Our results suggest that Turkey - Armenia may be the center of origin and diversity of *L. serriola* and may therefore have the greatest diversity of resistance genes.

Our characterization of the diversity of resistance genes and the genetic mechanisms generating it will allow informed exploration, *in situ* and *ex situ* conservation, and utilization of germplasm resources for disease control. The results of this project provide the basis for our future research work, which will lead to a detailed understanding of the evolution of resistance genes in plants.

Achievements

We comprehensively studied the genetic structure of wild lettuce, and genetic diversity of resistance genes in wild and cultivated accessions. This is one of the first analyses of such loci at the molecular level for any plant species. We documented a variety of genetic events occurring at clusters of resistance genes, including gene conversion and unequal crossing-over. We also demonstrated the surprising result that homologs from different species (orthologs) were more similar than homologs within a haplotype (paralogs). These studies complimented more reductionist molecular analyses of the major cluster being undertaken in the lab. Together, these studies resulted in the formulation of a new model for resistance gene evolution (Michelmore and Meyers, 1998). Our data suggested that, contrary to popular belief, resistance genes are evolving fairly slowly rather than resistance gene clusters being dynamic arrays with frequent sequence exchange between paralogs generating new resistance specificities.

Our findings have several basic and applied consequences. Our studies provide the basis for testing specific hypotheses concerning the relative importance of duplication, gene conversion, unequal crossing-over and divergent selection in the evolution disease resistance genes. Our characterization of the diversity of resistance genes and the genetic mechanisms generating it will allow informed exploration, *in situ* and *ex situ* conservation, and utilization of germplasm resources for disease control. The diversity of haplotypes indicates a rich source of resistance genes. However, the slow evolution of resistance genes suggests that there may be only a finite number of resistance genes against a particular pathogen available within the sexually compatible gene pool and that the same resistance gene could be introgressed multiple times from wild species. We now have evidence of this happening multiple times in lettuce breeding.

There was a regular transfer of technology, expertise, and materials between the Institute of Evolution and University of California, Davis. Two postdocs (D. Sicard and H. Kuang) spent considerable time at both institutions. AFLP technology was transferred to the Institute of Evolution where it is now routinely used for numerous projects. An extensive and unique set of *Lactuca* germplasm was collected. This was characterized in both locations as well as transferred to UC Davis for long-term maintenance. These characterized populations form the basis of the next phase of our research.

Final Report

Introduction

Proper use of wild genetic resources requires a good understanding of genetic structure, diversity, divergence and ecological correlates of the wild species. However, little is known in these aspects for *L. serriola* L. ($2n=18$), the progenitor of cultivated lettuce, *L. sativa* (Lindqvist 1960; Zohary 1991, Kesseli *et al.*, 1991; Hill *et al.*, 1996). These wild and cultivated ~~species~~ are cross-compatible and interfertile. Lettuce breeders have made extensive use of this ~~wild~~ species mainly for disease resistance (Crute, 1990).

Genetic resistance is often the least expensive, most effective, and ecologically sound method of disease control. An important use of germplasm in crop improvement is as a source of disease resistance (Leppik, 1970; Nevo *et al.*, 1985; Dale, 1991, Lenne and Wood, 1991). However, measuring diversity in the genetic basis of resistance has until recently been difficult. Typically, diversity of resistance in natural populations or germplasm collections has been analyzed by assessing resistance phenotypes (Jana and Nevo, 1991; Burdon, 1996, 1997). The recent cloning of resistance genes provides the opportunity to assay variation specifically at resistance loci using molecular markers derived from sequences of the resistance ~~genes~~.

Very little is known about the genetic mechanisms generating diversity at resistance loci. With the increasing number of resistance gene homologs (including both orthologs and paralogs) being cloned, it is now possible to investigate the relative significance of gene conversion, unequal crossing-over, point mutation and selection for divergence on the evolution of disease resistance genes in *Lactuca* spp. This will provide detailed information on the mechanisms of evolution of resistance genes in plants.

Disease resistance genes are located at least four major clusters in the lettuce genome (Kesseli *et al.*, 1994). Over eleven resistance specificities have been mapped to the major cluster. Molecular analysis revealed 24+ resistance gene paralogs at this locus (Meyers *et al.*, 1998a & b). One such paralog has recently been shown by mutation and transgenic complementation analysis to encode *Dm3* (Chin *et al.*, unpublished; Shen *et al.*, unpublished).

We made extensive collections of wild germplasm of *Lactuca* spp. as part of this BARD project. We studied the genetic structure of wild populations *L. serriola* using AFLP markers and analyzed the genetic diversity of resistance genes in Israel and California populations. We

documented a variety of genetic events, such as gene conversion and unequal crossing-over, occurring at clusters of resistance genes. The information and resources generated by this project are the foundation of our continuing studies on resistance gene evolution.

Materials and Methods

Collection of germplasm and materials used in this project

We made extensive collections of wild germplasm of *Lactuca* spp. in addition to the collection already available at UC Davis. We made 3,500 collections of *L. serriola* and 300 collections of *L. saligna* from the above regions and sampled most of the climatic and edaphic ranges of these species. Collections were made throughout the Near East and Europe as well as several collections from other parts of the world. We sampled most intensively in California, Israel, Jordan and several locations elsewhere in Europe as well as the Irano-Turanian regions where *L. serriola* may have originated (see below).

Genetic diversity of natural populations

Genetic structure of natural populations of *L. serriola* was studied using AFLP (Amplified Fragment Length Polymorphism) markers. AFLP was performed following Vos *et al.* (1995). Raw AFLP marker data were analyzed using Popgene (Yeh *et al.*, 1997) to calculate genetic distance (Nei, 1978), polymorphic loci (1% criterial), genetic diversity (Nei, 1973), and mean number of alleles. The 41 populations were divided into five groups (Europe, Israel, Turkey, Armenia, and Davis California), based on geographic locations. Dendrogram showing genetic relationship among populations was constructed using Nei's genetic distance (1978) with the method of UPGMA (StatSoft Inc., 1995).

Genetic diversity of resistance genes

A microsatellite marker, MSAT15-34, and two SCAR markers, ALRR and MLRR, were assayed for diversity of resistance genes. This microsatellite sequence is located in an intron of at least 5 members of the *RGC2* gene family in the cultivar Diana (Fig. 2 of Sicard *et al.*, 1999; Okubara *et al.*, 1997; Meyers *et al.*, 1998a). The two SCAR markers amplify regions encoding LRRs from multiple *RGC2* family members (Sicard *et al.*, 1999). These markers are duplicated in the cluster and amplify from multiple members of the *RGC2* family. Genotypes were therefore considered as haplotypes because it was not possible to distinguish between alleles of the same

gene and different paralogs. Haplotype diversity was estimated using the Shannon diversity index (Peet, 1974).

Detection of gene conversion

Nine resistance genes (RGC2) from *L. sativa* cv. Diana had been sequenced previously (Meyers *et al.*, 1998a). These nine genes were clustered over a 4.2 Mb region. These sequences were analyzed in greater detail using the GENECONV software (<http://lado.wustl.edu/~sawyer/geneconv/index.html>) to detect gene conversions among paralogs. GENECONV extends the methods of Sawyer (1989) in several ways (Sawyer, 1998). GENECONV uses only polymorphic sites in the alignment for analysis; monomorphic sites are ignored. Consequently, the effects of conserved regions and mutation hot spots are minimized. Two mismatch penalties (gscale=1 and 2) were chosen to run GENECONV. An event with $p < 0.01$ (simulated global p -value) was considered as positive. Each positive event when one gscale is used must be verified by the other gscale ($p < 0.05$). When the length or p value of the detected tract varied between the two gscale, the result with the smaller p value was chosen. If their boundaries varied significantly, DNA similarity along the tract and its flanking sequences were used to determine the conversion tract.

Cloning and sequencing orthologs and paralogs from six genotypes

Primers RLG3F2 and RLG3R2 (Shen *et al.*, 1998) were used to amplify a 700 bp region from the middle of RGC2 genes from six genotypes, two each from *L. sativa*, *L. serriola*, and *L. saligna*. The PCR products were cloned and then sequenced using an ABI377 automated DNA sequencer. Dendrogram of these sequences was constructed using UPGMA to show the relationships between paralogs and orthologs.

Results

Genetic parameters of natural populations

Genetic summary of all populations, based on all 319 AFLP markers, is shown in Table 1. The mean number of alleles per locus (A), percentage of polymorphic loci (P) and genetic diversity (He) for the whole population were 1.923, 92.3% and 0.185, respectively. According to the estimates of A (1.679) and P (67.9%), the most polymorphic group was the Turkish group. However, the Armenian group had the largest He (0.154). The inconsistency was probably due to

the difference of sampling sizes. The Israeli group had the lowest values for all three estimates, even though it had the largest sampling size.

Six of the 10 Turkish populations (Istanbul, Bolu, Kaiseri, Kaiseri-Sivas, Sivas and Ankara), three of five Armenian populations (Chor, Abovian and Artanish) and two of four Davis Californian populations (Yolo landfill and Road 29) had genetic diversity $He > 0.1$. The most polymorphic population was Abovian of Armenia, with $He = 0.132$. By contrast, all European and Israeli populations had genetic diversity $He < 0.1$. All European populations except Pavia, Koblenz and two Nancy populations had genetic diversity $He < 0.05$. These results suggested that Turkey and Armenia were (part of) the center of diversity of *L. serriola*. The least polymorphic populations were Shefayim of Israel, Wageningen of Netherlands, Landau of Germany, Antwerpen of Belgium, all with $He < 0.01$.

Genetic distance among populations

Unbiased genetic distance (Nei 1978) among populations was calculated (data not shown). The largest genetic distance was obtained between Tel Hanan of Israel and Alcanice of Spain ($D = 0.270$), and the smallest was between Tel Hanan and Haifa Freud Road ($D = 0.009$), both on Mt. Carmel, Israel, separated by about 15 km. Dendrogram for 41 populations was constructed using genetic distance with the method of UPGMA (Fig. 1). The populations from each region (Turkey, Armenia, California USA, Israel and Europe) were clustered together, with a few exceptions. The Israeli group was the least related to other groups, even though it was geographically close to Turkey. On the other hand, the California group, which was located in a different continent, was the most related to the Turkish populations. The Armenian group was the second closest to the Turkish group.

*Center of origin of *L. serriola**

The center of origin of a species should have most of the alleles that are present in its peripheral populations; in contrast, the peripheral populations may have only part of diversity present in the center of origin. We calculated the numbers of alleles in all loci that were present in one group but absent from another group (Table 2). These numbers are comparable, considering that each group (except Armenia) had a similar number of genotypes. Turkey had a large number of alleles that were not present in other groups. For example, 110 alleles were present in Turkey but not in Israel (Table 2). On the other hand, only small numbers of alleles

were present in other groups but absent in Turkey. For example, only 32 alleles were present in Israel but not found in Turkey. This data suggest that Turkey might be the center origin of *L. serriola*.

Polymorphism for the microsatellite MSAT15-34 among cultivated and wild accessions of Lactuca spp.

MSAT15-34 was assayed on a total of 148 genotypes from seven *Lactuca* spp. to determine the level of polymorphism at the major cluster of resistance genes. We analyzed the amplification products as combinations of array sizes rather than alleles. Both the size and number of fragments amplified with MSAT15-34 primers were highly variable. Twenty-seven different fragment sizes ranging from 66 to 125 base pairs were observed. The complexity of haplotype (the number of arrays per haplotype) varied from one to seven products (Fig. 4A of Sicard *et al.*, 1999).

Twenty-four haplotypes were detected in 74 accessions of the cultivated species, *L. sativa*. The major cluster contains genes *Dm1*, *Dm2*, *Dm3*, *Dm6*, *Dm14*, *Dm15*, *Dm16*, and *Dm18* (Farrara *et al.* 1987; Maisonneuve *et al.* 1994). Molecular marker haplotypes were highly correlated with *Dm* genes detected phenotypically at the major cluster but not with *Dm* genes in other clusters (Sicard *et al.*, 1999).

Fifty-one different haplotypes were detected in the 74 accessions of wild *Lactuca* germplasm that had previously been identified as being resistant to all known isolates of *B. lactucae*. The majority of accessions had unique haplotypes. The most common haplotype had a frequency of less than 0.08. Therefore there was very little redundancy amongst these lines and they will likely be as a rich source of new resistance genes (Sicard *et al.*, 1999).

Polymorphism for the MSAT15-34 among populations of L. serriola from Israel and California

To investigate the distribution of polymorphism in the major cluster in natural populations, MSAT15-34 was analyzed in individuals collected from 13 populations in Israel and California. Haplotypes were determined successfully for 492 out of 505 samples. Fourteen array sizes were observed in these populations; all except three had previously been observed in the resistant *L. serriola* germplasm (Table 2 and Fig. 4C of Sicard *et al.*, 1999). In contrast, out of a total of 33 haplotypes identified, only eight had been observed previously in the resistant germplasm. Haplotypes containing a single array were always the most frequent within a country as well as within a population. Six arrays were detected in one accession indicating that

as much duplication of *RGC2* genes can occur in natural populations of *L. serriola* as in *L. sativa* (Meyers *et al.*, 1998a).

Levels of diversity were compared between Israel and California. In Israel, 17 different haplotypes were detected and in California, 22 (Table 4A of Sicard *et al.*, 1999). In each country, one haplotype was prevalent and all the others had frequencies less than 0.15 (Table 4B of Sicard *et al.*, 1999). Only six haplotypes were common to samples from both Israeli and Californian (Table 4B of Sicard *et al.*, 1999). Moreover the most frequent Israeli haplotype was never observed in California and the most frequent haplotype in California was only represented by two Israeli individuals in a single population. The diversity within each country was estimated using the Shannon index. This index takes into account the number and the frequency of haplotypes. The Californian population was as polymorphic as the Israeli population (Table 4A of Sicard *et al.*, 1999).

Fingerprints of the LRR encoding regions and correlation between markers

To provide additional information on polymorphism in *RGC* genes, two pairs of primers were designed to amplify from multiple *RGC2* sequences encoding two different regions of the LRR (Fig. 2 of Sicard *et al.*, 1999). All samples of cultivated and wild germplasm were analyzed (Table 2 of Sicard *et al.*, 1999). Approximately ten samples in each population of *L. serriola* representing the diversity of MSAT15-34 haplotypes were also studied (Table 4 of Sicard *et al.*, 1999). A product of 703 bp was amplified from sequences encoding the anterior LRR region (ALRR) and a product of 661 bp from the sequences encoding the mid-LRR region (MLRR). Surprisingly, no length polymorphism was detected with either set of primers from 150 germplasm and 137 *L. serriola* samples (Fig. 6A of Sicard *et al.*, 1999). Therefore there is minimal variation in the numbers of LRRs in these regions for multiple copies of the gene across diverse germplasm.

Polymorphism was detected after digestion of the amplification products with *Tsp509I*, an endonuclease with a four base pair recognition site. After digestion of the ALRR product, 46 different haplotypes were obtained in the germplasm collection. After digestion of the MLRR product, 63 haplotypes were detected. Sixteen and 15 haplotypes were observed in natural populations of *L. serriola* in California and Israel respectively (Table 5 of Sicard *et al.*, 1999). As expected, the sum of fragment sizes observed exceeded the size of the undigested band, reflecting amplification from multiple *RGC2* genes. The diversity of *RGC2* haplotypes correlated

well with the MSAT15-34 haplotypes within each species (Table 3 of Sicard *et al.*, 1999). The levels of diversity detected by the three markers were not significantly different using the Shannon index (Table 5 of Sicard *et al.*, 1999). The level of resolution with the *RGC2* fingerprints was slightly lower than with MSAT15-34, i.e. some groups with identical *RGC2* fingerprints could be split on the basis of MSAT15-34 data. Less frequently, MSAT15-34 haplotypes could be split on the basis of *RGC2* SCAR markers. Therefore, MSAT15-34 was the most informative marker for diversity in *RGC2* sequences.

More variation between paralogs than orthologs in the middle of RGC2 genes

To provide a detailed analysis of inter-genotypic variation, we compared the sequences in the anterior region of the LRR of *Dm3* paralogs in a total of six genotypes of *Lactuca* spp. A 700-750bp fragment was amplified from this region using PCR from two genotypes each of *L. sativa*, *L. serriola*, and *L. saligna*. A total of 277 PCR products were cloned and sequenced. Ninety-two sequences were identified, 35 of which were found more than twice and analyzed further. Phylogenetic tree of these 35 clones was constructed (Fig. 2). Phylogenetic analysis demonstrated much greater similarity between than within genotypes, indicating that resistance genes were evolving fairly slowly and paralogs were not being homogenized by gene conversion or unequal cross-over events. This in part led to the development of new ideas as to how resistance genes evolve (Michelmore and Meyers, 1998). We are currently extending this analysis to other regions of the gene, particularly the region encoding the C-terminal LRR that may be involved in determining specificity.

Most gene conversions were in 3' half of RGC2 genes

Using Sawyer's software, GENECONV, we investigated gene conversions among nine *Dm3* homologs in the cultivar Diana. Twenty-two gene conversions and five intergenic recombinations were identified (Table 3, 4). The length of gene conversion tracts varied from 72-2413bp, mostly in the range of 100-250bp. Eighteen gene conversions and four recombinations occurred in the 3' half of the genes, where sequences were more conserved than in the 5' half. The frequent gene conversion created a patchwork structure for some genes (Fig. 3).

Discussion

We developed highly informative molecular markers for assessing resistance gene diversity from the sequences of resistance gene homologs. Using these markers, high level of haplotype diversity was observed in both wild and cultivated species, suggesting that the diversity of resistance phenotypes previously identified within *L. sativa* is not an artifact of selection for resistance by plant breeders. The large number of different haplotypes detected in wild germplasm indicated that there is little redundancy and that these accessions will be a rich source of new resistance genes.

Although a large number of array sizes were observed, the high correlation between MSAT15-34 and the two SCAR markers suggests that none of them are evolving rapidly enough to break down the correlation. It also indicates that although unequal crossing-over is occurring to change copy number (as evidenced by the changes in numbers of arrays), the cross-over events are not occurring frequently within the genes.

Very little length variation was observed in either the ALRR and MLRR markers amplified from multiple paralogs in diverse germplasm spanning four species. These were of the 703 and 661 bp regions encoding part of the LRR region. This was surprising as changes in the number of LRRs has been proposed to be involved in the evolution of new resistance specificities (Jones and Jones 1997; Hammond Kosack and Jones 1997). In contrast, restriction site polymorphisms were detected suggesting that substitution or small indels are more frequent than large insertions or deletions. This finding extends data obtained from the sequencing of nine RGC2 paralogs from cv. Diana; there was little length variation in the seven exons (Meyer et al. 1998b). However, within each LRR, one motif was hyperpolymorphic suggesting that point mutation was probably an important mechanism of variation at resistance gene clusters. Together these results suggest that point mutations are more important than variation in the number of LRRs in this region to the generation of diversity within and between species.

The relationship of orthologs and paralogs in 700bp region in the middle of RGC2 genes showed that there is much greater similarity between than within genotypes. This may indicate that resistance genes were evolving fairly slowly and paralogs were not being homogenized by gene conversion or unequal cross-over events. The genetic data on the instability of resistance gene clusters in several species had previously led to the idea that such clusters are inherently unstable and that frequent rearrangements resulted in the rapid evolution of new resistance specificities. However, our data suggests that resistance genes are not evolving rapidly and that

while there is some unequal crossing-over or gene conversion occurring, it is rare and insufficient to homogenize the sequences within the multigene family. This conclusion is based on the information at the 5' half of the RGC2 genes. We are currently sequencing other regions of the gene, particularly the region encoding the C-terminal LRR that may be involved in determining specificity. Sequence comparison among nine genes showed that gene conversion and unequal crossing-over is more frequent in the C-terminal than in the N-terminal region, and patchwork structure was present the 3' half of some RGC2 genes. However, the gene conversion was not frequent enough to homogenize the region.

The evolutionary mechanisms generating diversity in resistance genes in plants remains unclear. Our studies contrast with conclusions drawn from data for the *Cf4/9* locus in tomato and the *RPP5* locus in *Arabidopsis* that suggest new resistance specificities are frequently generated by sequence exchange between paralogs (Parniske *et al.*, 1997; Noel *et al.*, 1999). Our data, however, agree with the studies on *RPM1* in *Arabidopsis* that concluded that resistance genes are evolving slowly (Stahl *et al.*, 1999). Our BARD project documented a range of genetic events occurring at the major cluster of resistance genes and provided the genetic resources for our current efforts that are focused on determining the relative importance of duplication, gene conversion, unequal crossing-over and divergent selection in the evolution disease resistance genes in natural populations of *L. serriola*.

References

- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999). Dynamics of disease resistance polymorphism at the *RPM1* locus of *Arabidopsis*. *Nature* 400:667-671.
- Burdon JJ (1996). The role of race specific resistance in natural plant populations. *Oikos* 76: 411-416.
- Burdon JJ (1997). The evolution of gene-for-gene interactions in natural pathosystems. In: Crute IR, Holub EB, Burdon JJ (eds) *The gene-for-gene relationship in plant-parasite interactions*. CAB international, Oxon, UK.
- Crute IR (1990). Resistance to *Bremia lactucae* (downy mildew) in British populations of *Lactuca serriola* (prickly lettuce). In: Burdon JJ, Leather SR (eds) *Pests, pathogens and Plant communities*. Blackwell Scientific Publications, Oxford, UK, pp 203-217.
- Dale MFB (1991). The use of genetic resources in breeding and breeding research. In: Hawkes JG (eds) *Genetic conservation of world crop plants*. Academic Press, London, UK, pp 3-10.
- Farrara B, Illott TW, Michelmore RW (1987). Genetic analysis of factors for resistance to downy mildew (*Bremia lactucae*) in lettuce (*Lactuca sativa*). *Plant Pathol* 36: 499-514.
- Hammond-Kosack K, Jones JDG (1997). Plant disease resistance genes. *Annu Rev Plant Physiol Plant Mol Biol* 48: 575-607.

- Hill M, Witsenboer H, Zabeau M, Vos, P, Kesseli R, Micheltmore RW (1996). AFLP fingerprinting as a tool for studying genetic relationships in *Lactuca* spp. *Theor Appl Genet* 93: 1202-1210.
- Hutcheson K (1970). A test for comparing diversities based on the Shannon formula. *J Theor Biol* 29: 151-154.
- Jana S, Nevo E (1991). Variation in response to infection with *Erysiphe graminis hordei* and *Puccinia hordei* in some wild barley populations in a centre of diversity. *Euphytica* 57: 130-140.
- Jones DA, Jones JDG (1997). The roles of leucine-rich proteins in plant defences. *Adv Plant Pathol* 24: 89-167.
- Kesseli RV, Ochoa O, Micheltmore RW (1991). Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce. *Genome* 34: 430-436.
- Kesseli RV, Paran I, and Micheltmore RW (1994). Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics* 136:1435-1446.
- Lenne JM, Wood D (1991). Plant diseases and the use of wild germplasm. *Annu Rev Phytopathol* 29: 35-63.
- Leppik EE (1970). Gene centers of plants as sources of disease resistance. *Annu Rev Phytopathol* 8: 323-344.
- Lindqvist K (1960). On the origin of cultivated lettuce. *Hereditas* 46: 319-350.
- Maisonneuve B, Bellec Y, Anderson P, Micheltmore RW (1994). Rapid mapping of two genes for resistance to downy mildew from *Lactuca serriola* to existing clusters of resistance genes. *Theor Appl Genet* 89: 96-104.
- Meyers BC, Chin DB, Shen KA, Sivaramakrishnan S, Lavelle DO, Zhang Z, Micheltmore RW (1998a). The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10:1817-1832.
- Meyers BC, Shen KA, Rohani P, Gaut BS, Micheltmore RW (1998b). Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell* 10:1833-1846.
- Micheltmore RW, Meyers BC (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* 8:1113-1130.
- Nei M (1973). Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321-3323.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.
- Noel L, Moores TL, van der Biezen EA, Parniske M, Daniels MJ, Parker JE, Jones JD (1999). Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of *Arabidopsis*. *Plant Cell* 11: 2099-2112.
- Nevo E, Moseman JG, Beiles A, Zohari D (1985). Patterns of resistance of Israeli wild emmer wheat to pathogens: predictive method by ecology and allozyme genotypes for powdery mildew and leaf rust. *Genetica* 67: 209-222.
- Okubara PA, Arroyo-Garcia R, Shen KA, Mazier M, Meyers BC, Ochoa OE, Kim S, Yang C-H, Micheltmore RW (1997). A transgenic mutant of *Lactuca sativa* (Lettuce) with a T-DNA tightly linked to loss of downy mildew resistance. *Mol Plant-Microbe Interact* 10: 970-977.
- Parniske M, Hammond-Kosac KE, Golstein C, Thomas CM, Jones DA, Harrison K, Wulff BB, Jones JD (1997). Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* 91: 821-832.

- Peet RK (1974). The measurement of species diversity. *Ann Rev Ecol System* 5: 285-307.
- Sawyer, S. A. (1989). Statistical tests for detecting gene conversion. *Molec. Biol. Evol.* 6, 526-538.
- Sawyer, S. A. (1998). GENECONV: A computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis, available at <http://www.math.wustl.edu/~sawyer>.
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999). Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400: 667-671.
- Yeh FC, Boyle T, Yang R-C (1997). Popgene version 1.21. Microsoft Window-based Freeware for population genetic analysis. Quick user guide. <http://www.ualberta.ca/~fyeh>.
- Zohary D (1991). The wild genetic resources of cultivated lettuce (*Lactuca sativa* L.). *Euphytica* 53: 31-35.
- Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21): 4407-4414.

Publications:

- Kuang, H., Michelmore, W. R., van Eck, H., Nevo, E. Genetic diversity of wild lettuce, *Lactuca serriola*: Center of origin and diversity, old and new colonizations. (In preparation^a).
- Kuang, H., Nevo, E., Michelmore, W. R. Gene conversion between paralogs of *RGC2* genes in *L. sativa* cv. Diana. (In preparation^b).
- Michelmore, R.W. and Meyers, B.C. (1998). Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* 8:1113-1130.
- Sicard, D., Woo, W.-W., Arroyo-Garcia, R., Ochoa. O., Korol, A., Nevo, E., and Michelmore, R.W. (1999). Molecular diversity at the major cluster of disease resistance genes in cultivated and wild *Lactuca* spp. *Theor. Appl. Genet.* 99:405-418. (In Press. Attached).
- Woo, S.-S., Nevo, E., and Michelmore, R.W. Comparisons of orthologs and paralogs of resistance genes reveals greater similarity between species than within haplotypes. (In prep.)

APPENDIX

Fig 1. Dendrogram showing the genetic distance among natural populations of *L. serriola*. 315 AFLP markers were used to calculate Nei's genetic distance matrix of natural populations. Dendrogram was constructed using the method of UPGMA.

Fig 2. Phylogenic tree of RGC2 gene family sequences from six genotypes. The anterior region of LRR was amplified by PCR with RLG3F2 and RLG3R2 primers. The PCR products (700-750bp in length) were cloned and a total number of 277 clones were sequenced. The numbers in parenthesis represent the number of identical sequences found in the total sequences analyzed. Among 92 unique sequences, 35 sequences were found more than twice.

Fig 3. Diagram showing the chimerical structure of RGC2J, which shares identical DNA sequences to different genes in different regions. This structure was most likely created by several gene conversions between RGC2J and other genes. Polymorphic sites of each gene conversion tract are shown. The numbers in the top line indicate the position of each gene conversion tract in the alignment.

Figure 1

Figure 2

	300	3600	4220	4240	4320	4460	4480
RGCC2J :							
RGCC2A :							
RGCC2N :							
RGCC2D :							
RGCC2B :							
RGCC2C :							
RGCC2S :							
RGCC2O :							
RGCC2K :							

Figure 3

Table 1. Genetic variation of 41 natural populations

Population	Collected Year	Sample size	Mean number of alleles per locus (<i>A</i>)	Polymorphic loci		<i>He</i>
				No.	% (<i>P</i>)	
TURKEY						
Goreme	98	20	1.312	100	30.86	0.081
Istanbul	98	20	1.312	100	30.86	0.109
Bolu	98	20	1.361	114	35.19	0.103
Ankara	98	10	1.328	106	32.72	0.113
Yakakent	98	19	1.276	89	27.47	0.052
Bergama	98	21	1.155	49	15.12	0.048
Kaiseri	98	20	1.399	128	39.51	0.120
Kaiseri-Sivas	98	19	1.343	110	33.95	0.102
Sivas	98	20	1.344	111	34.26	0.101
Afyon	98	20	1.228	72	22.22	0.070
Average		18.9	1.306	97.9	30.22	0.090
AEMENIA						
Chor	98	20	1.323	102	31.48	0.104
Abovian	98	20	1.408	130	40.12	0.132
Areni	98	20	1.296	81	25.00	0.080
Sevan	98	16	1.288	94	29.01	0.098
Artanish	98	18	1.387	123	37.96	0.125
Average		18.8	1.340	106	32.71	0.108
Israel						
N. Narazeth	94	13	1.217	70	21.60	0.051
S. Narazeth	94	32	1.155	50	15.43	0.034
Haifa Freud Rd	94	48	1.226	73	22.53	0.045
Shefayim	94	24	1.022	7	2.16	0.004
Beer Sheva	94	39	1.189	61	18.83	0.056
Tel Hanan	94	24	1.189	61	18.83	0.043
Netaniyya	94	21	1.228	72	22.22	0.089

Haifa Oren Rd	94	23	1.186	60	18.52	0.035
Average		28	1.176	56.8	17.52	0.045

Europe

Salamanca	94	9	1.150	47	14.51	0.041
Alcanice	94	6	1.083	26	8.02	0.028
Pyrenees	97	12	1.171	51	15.74	0.041
Aix le Bains	97	18	1.171	51	15.74	0.049
Mont Blanc	97	16	1.098	31	9.57	0.027
Nancy 1	97	9	1.220	69	21.30	0.084
Nancy 2	97	9	1.169	52	16.05	0.072
Pavia	97	16	1.296	91	28.09	0.095
Luzern	97	19	1.105	33	10.19	0.019
Landau	96	17	1.038	12	3.70	0.006
Koblenz	96	14	1.213	67	20.68	0.057
Rostock	96	7	1.102	32	9.88	0.030
Antwerpen	97	13	1.010	3	0.93	0.003
Wageningen	98	25	1.114	36	11.11	0.010
Average		13.6	1.139	42.9	13.25	0.040

California

Yolo Landfill	94	55	1.444	144	44.44	0.112
Putah Creek		31	1.325	105	32.41	0.087
Road 29		51	1.486	157	48.46	0.117
Road 32	95	29	1.356	115	35.49	0.091
Average		41.5	1.403	130.3	40.2	0.102

Table 1 shows the genetic diversity of 41 natural populations as detected by AFLP markers. For AFLP markers, each locus has two alleles, presence and absence of an amplified fragment. If a population is homozygous at one locus, it has 1 allele, otherwise 2 alleles. "Mean number of alleles" measures the average of the number of alleles in all screened loci. He is genetic diversity.

Table 2. Number of alleles detected in one group but absent in another

A.

Absent in	Present in				
	Israel	Davis	Europe	Armenia	Turkey
Israel	0	69	69	84	110
Davis		0	46	62	82
Europe			0	68	89
Armenia				0	75
Turkey					0

B.

Present in	Absent in				
	Israel	Davis	Europe	Armenia	Turkey
Israel	0	36	59	54	32
Davis		0	69	65	37
Europe			0	48	21
Armenia				0	27
Turkey					0

Table 3. Gene conversions among nine *Dm3* paralogs

No.	Sequence ¹	Tract ²	Length	<i>P</i> (<)
1	(B, S)-(C, D) ³	1428-3840	2413	0.0001
2	D-outer	3128-3508	381	0.0001
3	S-outer	1495-1661	167	0.0001
4	J-A	3466-3687	222	0.0001
5	J-D	3935-4054	120	0.0062
6	J-N	4064-4304	241	0.0001
7	J-D	4267-4388	122	0.0014
8	J-A	4411-4533	123	0.0002
9	A-B	3734-4196	463	0.0001
10	O-(B, S)	961-1493	533	0.0001
11	N-C	4027-4543	517	0.0001
12	J-K	1894-2074	181	0.0006
13	A-S	3466-3840	375	0.0001
14	N-D	4316-4463	148	0.0069
15	J-B	5785-5897	113	0.0006
16	N-B	4322-4560	239	0.0001
17	A-S	6019-6272	254	0.0001
18	N-D	6022-6122	101	0.0002
19	J-D	6010-6100	91	0.001
20	N-S	5929-6109	181	0.0064
21	A-D	6022-6124	103	0.0065
22	J-C	6029-6100	72	0.0006

1. The letters represent RGC2 genes, for example "D" representing RGC2D. One of the two sequences (separated by "-") converted the other.
2. Distance from the first base of alignment.
3. The common ancestor of the two sequences in bracket.

Table 3 shows the gene conversions detected among nine RGC2 genes using software GENECONV. Most of the gene conversions were short tract and the majority were located in the 3' half of the genes (after 3000th bp in the alignment).

Table 4. Intergenic recombination ($p < 0.0001$)

Event	Sequences	Tract	Tract length (>)
1	A-N	1-3768	3768
2	B-S	1-3583	3583
3	C-D	1-3940	3940
4	J-(B, S)	1-369	369
5	A-C	6024-6378	355

Note: See Table 2 for the meaning of each column.

Table 4 shows the intergenic recombinations detected by GENECONV. Like gene conversion, intergenic recombination results in a segment with higher than expected similarity in two otherwise diverse DNA sequences. Unlike gene conversions, intergenic recombination created a long shared tract which extends to the end of the sequences analyzed.