Canola Agronomic Research Program

Project AG#2002-25

Final Report

Assessing Genetic Diversity of Lygus Pest Species In Crop and Non-Crop Habitats

R. G. Foottit

Principal Investigator

ABSTRACT

This project investigated the development and use of DNA microsatellite flanking region sequences for the analysis of genetic diversity in *Lygus* species both from the perspective of species relationships and from the population-level characteristics of certain pest *Lygus* species, particularly the Tarnished Plant Bug , *Lygus lineolaris*. Methods were developed or refined for the establishment of informative sequencing loci which would provide combined information on both within- and between species patterns of variation. Analysis of interspecific patterns confirmed the general classification of *Lygus* based on morphology, however, molecular data established the re-positioning within the phylogenetic tree of *Lygus borealis* and *Lygus keltoni*, both economically important species. Analysis of genotypic patterns within *Lygus lineolaris*, based on extensive geographic and host plant based sampling, uncovered the presence of two major genotypes. These genotypes often were sympatrically associated although they appear to be reproductively isolated.

Impact

The re-positioning of *Lygus borealis* and *Lygus keltoni* has implications for the management of and the future research on the genus *Lygus*. An accurate phylogenetic tree, indicating the position and relationships among species and species groups is essential for non-target testing of biological control agents and conversely for the testing of host range, host suitability and effectiveness of control agents. The presence of major genotypic differences among populations of *Lygus lineolaris* has implications for the future management of this pest. The current situation as revealed by the DNA sequence analysis may reflect a dynamic, on-going process of selection and adaptation to changing agricultural practices and environmental modification.

1.0 Introduction

Several *Lygus* species are an important pest of several crops in North America, particularly of canola and seed alfalfa. The principal pest species in Canada is *Lygus lineolaris*. In western Canada, *L. borealis* is important on canola, and *L. keltoni* is becoming more significant in Alberta and Saskatchewan.

Recently, this genus has been the subject of biological control work using species of braconid (*Peristenus*) parasites. It is becoming increasingly clear that attempts at biological control of insect pests will require detailed knowledge of the demography and genetic composition of populations and species (Roderick 1996). The ability to predict population sensitivity to biological control organisms is essential to the planning, modelling and successful release of agents. The population structure of *L. lineolaris* across its range is presently unknown. Successful implementation of biological control programs for *L. lineolaris* will need estimates of the extent of genetic divergence within and between populations. Molecular markers have proven to be a valuable tool in the identification and phylogenetic analysis of *Lygus* species (Schwartz and Foottit 1997). These same general technical approaches can be used to analyze the structure and diversity of populations and to serve as identifying markers for routine monitoring in the field.

Microsatellites are tandemly repeated short sequences of DNA, between one and five base pairs, that evolve rapidly and are particularly useful as sources of measurable genetic variation, particularly when other methods show insufficient variability (Goldstein and Schlotterer 1999, Queller et al.1993). They are dispersed widely throughout eukaryote genomes (Tautz 1989) and they are highly polymorphic, usually having more than a dozen alleles at a single locus. Microsatellites are widely used in the study of natural populations because many loci can be analyzed and scored and they provide independent genealogies for population studies. The total size of microsatellite arrays is small, generally less than 300 base pairs, allowing a single locus to be PCR amplified.. However, microsatellite length variants are subject to convergence and each locus provides only a single character. In addition, because sequence adjacent to the repeat track is often quite variable, primers developed for amplifying a microsatellite in one species often fail to work in other species, or amplify non-homologous regions. Thus, while providing an excellent means of characterizing a sample or population, they are difficult to analyse phylogenetically. We have pursued an approach which links phylogenetically informative sequence data to microsatellite track length data by developing primers to amplify the regions immediately flanking the repeat track. These flanking region sequences show various degrees of variability, from quite variable within species, thus giving good population level information, to rather stable within species, providing good diagnostic characters (unpublished data).

In this study, we investigated the development and use of microsatellite flanking regions to uncover species level relationships and population-level characteristics within pest *Lygus* species.

2.0 Methods

DNA was extracted from the abdomen of individual lygus bugs using a modified CTAB extraction method, and the remainder of the body retained as a voucher. DNA loci both crossing and adjacent to tandem dinucleotide repeats (microsatellites) were developed. See Appendix A for a full description of methods extraction, library construction and primer development).

Sequences were determined for three microsatellite flanking region loci (LL13F, LL42F and LL56F) and for one locus containing a short microsatellite track (LL34). Primer sequences are listed in Table 1.

PCR amplifications were performed using Qiagen Taq PCR Master Mix in a Techne Genius thermocycler under the conditions specified in Table 1.

Sequencing of survey samples was done using direct bidirectional sequencing of PCR product on a LI-COR/NEN Global IR^2 automated sequencing system with DYE-namic Sequencing Kit (Amersham) using labelled custom primers (sequences for internal primers given in Table 2). Cycle-sequencing reactions were done on a Techne Genius thermocycler with a program of 30 cycles of 95°C for 30 s + 52°C for 15 s + 70°C for 40 s.

3.0 Material

Samples of *Lygus lineolaris* were collected in throughout southern Ontario and in the Gatineau and Pontiac regions of Québec in 1999, 2000, 2001, and 2003. Collections of *L. lineolaris* and other *Lygus* species (primarily *L. keltoni* and *L. borealis*) were taken at 142 localities in Alberta, Saskatchewan and Manitoba in 2002 and 2003. Continental transects (Pennsylvania to Colorado and northern Ontario to British Columbia) of *L lineolaris* consisting of a total of 50 collections were made in 1999, and 80 samples *L. lineolaris* as well as several collections of *L. keltoni* and *L. borealis* were taken in 1995 throughout the agricultural areas of the Prairie provinces. Collecting sites are mapped in Figure 1.

Representative samples of other species were provided by M.D. Schwartz.

Specimens collected in the Prairie provinces in 2002 and 2003 were preserved in 95% ethanol. Other samples were frozen in the field in a liquid nitrogen dry-shipper and subsequently stored at -75 to -85°C.

4.0 Results

The number of samples for species other than *Lygus lineolaris* was limited. Thus analysis of geographic and host-related patterns has to date been restricted to *L. lineolaris*.

4.1 LL13F

Table 3 gives a reference sequence for LL13F based on majority consensus for samples of *Lygus lineolaris* 'group B'. All subsequent discussion is in terms of position relative to this reference sequence.

The majority of locus LL13F appears to be a protein-coding region, a reading frame beginning at position 2 resulting in no stop codons until position 584-586, and a second immediately following. The first stop is modified to glutamine in *L. keltoni* (homozygous, or heterozygous combination with a stop codon), and to lysine in some *L. lineolaris*.

4.1.1 Interspecific variation

Table 4 indicates states of sites which showed variation among species. Positions which vary within a species but which are in general otherwise constant across species are not included in this table. Most species are currently represented by a single sample and the indicated sequence differences therefore may not in general be taken as diagnostic. This locus also strongly supports the separation of *Lygus* from *Henrilygus* and *Nonlygus* (19 characters, data not shown) as established by Schwartz and Foottit (1998).

Specimens assigned to *Lygus lineolaris* fall into two distinct sequence patterns ('group A' and 'group B', see below) and are treated separately in Table 4.

The most striking feature of Table 4 is the consistent heterozygosity (dominated by G/A transitions) shared among several species at 19 positions. Because the apparent heterozygosity is present in sequence primed in both forward and reverse directions, it is believed that this is a real phenomenon, not a PCR artefact. However, relative signal strength of the contributing allelic states is usually not equal, but, within a taxon, the inequality is generally relatively consistent (Table 5). *L. keltoni*, *L. shulli* and *L. hesperus* show a similar trend at a significant subset of these sites. Since a 1:1 ratio would be expected for a diploid individual heterozygous for a single-copy gene, it is suggest that the observed patterns may be explained by assuming that there are additional copies of the gene either through gene duplication within a haplotype or by polyploidy. Within the degree of error of interpretation of chromatogram peak intensities, the observed frequencies may be reasonably interpreted as a 4-copy system.

All of the species demonstrating this pattern are western montane, except *L. solidaginis* (northern Great Plains) and *L. shulli* (primarily western montane, but extending across the content in the boreal zone).

4.1.2 Intraspecific variation

4.1.2.1 *Lygus lineolaris*

Two distinct groups of *L. lineolaris* may be defined based on differences at 14 positions (Table 4). Only one of these changes (position 260) results in an amino acid substitution (asparagines for praline). Group A occurs in southern Ontario and was also collected at several localities between Indiana and eastern Nebraska. In south-western Ontario south and west of Wellington County it is the predominant form. In eastern Ontario it is common from eastern Northumberland County to Frontenac County south of Highway 7 (approximate southern boundary of the Precambrian Shield), and less frequent eastward to the Quebec border. It has not been collected elsewhere in Ontario, other than a single individual near Calabogie in Renfrew County. Group B is present throughout the sampled areas of the continent.

At sites yielding individuals belonging to group A, group B individuals often (in eastern Ontario, almost always) occur. Despite the frequent co-occurrence at a site, there is no evidence of interbreeding between the two groups (no extensive heterozygosity at the relevant positions), except for one specimen heterozygous at several of these positions at one end of the sequenced fragment.

Most of the sequence variation among individuals of group A is unique to a single sample. However, variants of a G/A transition at position are shared among several collections, as are at

positions 179 (G/T transversion, T appearing only as heterozygote K), 250 (G/C transversion, C mainly as heterozygote S) and 551 (G/T transversion, T both homozygous and as heterozygous K). No geographic pattern is exhibited.

Within group B, a large proportion of individuals (30%) distributed through the sampled geographic range have identical sequence for LL13F.

Variants shared among more than three samples include single base polymorphisms at positions 49 (G to T), 55 (T to C), 65 (C to G, rarely A or T), 385 (T to G), 412 (T to C), 415 (T to A), 599 (A to T) and 615 (A to G) of the reference sequence, and a 3 base insertion after position 580. Frequencies for the variants are given in Table 7.

Alternate forms at positions 49, 55, and 599 as well as the insertion at base 580 each appear in about 20% of observed haplotypes in all adequately sampled geographic regions (Canadian prairies, southwestern Ontario, southeastern Ontario, West Quebec), and are broadly distributed within those regions. Occurrence of G at position 615 is also broadly distributed in all regions: the lower observed frequency is probably an artefact of the fact that individuals heterozygous for the insertion at position were not scored at this position (see below), biasing the ratio in favour of the more abundant variant.

Polymorphism at positions 65, 385 and 415 show some geographic patterns.

The A to T transversion at position 415 is present in about 5% of southern Ontario samples (both southwest and southeast); otherwise it appears only as a single heterozygous individual in West Quebec and another in Alberta.

Replacement of C with G at position 65 is almost entirely restricted to Renfrew County in Ontario, adjacent RMC Pontiac in Quebec, and three locations in central Alberta; exceptions include single individuals at Guelph (heterozygous G/T) and at Silver Lake in Frontenac Co. (heterozygous G/C). Replacement of C with A was also observed in Renfrew and Pontiac counties (one sample in each), and T occurs at this position only at Guelph (heterozygous G/T, same individual as above), one sample south of Ottawa (heterozygous C/T) and one specimen in Outaouais region of Quebec (homozygous).

At position 385, replacement of T with G is virtually absent in the prairies (1 sample in southern Manitoba), but rather broadly distributed in southern Ontario and west Quebec (present in 8 to 16% of haplotypes in these regions). However it was not found in the rather heavily sampled extreme eastern Ontario (Ottawa, Prescott-Russell, and Stormont-Dundas-Glengary).

Change of T to C at position 412 is broadly distributed throughout all regions, but shows marked differences in frequency: on the prairies, C occurs at this position in only 20% of haplotypes, but is progressively more abundant moving eastwards, ranging from 24% in southwestern Ontario to 54% in west Quebec.

All individuals homozygous for the insertion have the combination T-G at positions 599 and 615 respectively, while only one individual homozygous for no insertion has this combination, and no individual with G at position 615 has A at position 599 (although the converse is not true). Because the reverse sequencing primer for this locus is upstream of the indel site, this portion of the fragment was sequenced in one direction only, so that the sequence for individuals heterozygous for the indel was unreadable past that point. Thus, the status of these positions in those individuals is unknown.

There is no other indication of linkage among the polymorphic sites.

Within the apparent coding segment, 5 of the 6 point polymorphisms correspond to third position changes with respect to amino acid codons, but changes at position 65 are first position changes resulting in replacement of glutamic acid with glutamine or lysine. Note also that the 3-base insertion maintains reading frame integrity.

Two individuals, one from eastern Ontario (Silver Lake), the other from the Niagara Region, share a major deletion as well as two point mutations: G replaces A at position 138, T replaces C at position 190, and 202 bases (positions 238 to 439, almost one third of the entire fragment) have been deleted. This deletion violates reading frame integrity. The Niagara sample also has a insertion of A at position 204; the Silver Lake specimen has a 19 base insertion at position 580 (the same position as the 3-base insertion of other individuals).

Numerous samples exhibit unique variants, a few others are shared between two individuals. All of these were silent changes with respect to amino acid translation Populations from the Pontiac/Renfrew region of west Quebec and eastern Ontario are considerably more polymorphic than those of any other comparable area, both within and among collecting localities. About one third (20 of 56) variants present in one or two individuals were found in this region.

No clear correlation of any genotype with host plant was apparent, except that Brassicaceae are fairly rare as hosts for group A samples, or group B samples with bases other than C at position 65. This, however, is completely confounded with geography, since collections from cultivated *Brassica* species were common in the prairies, while collections from *Potentilla* and Compositae were more frequent in eastern Ontario and west Quebec.

No trends are apparent in year to year comparisons.

4.1.2.2 *Lygus borealis*

Variation among *Lygus borealis* samples is summarized in Table 8.

4.1.2.3 *Lygus keltoni*

Variation among *Lygus keltoni* samples is summarized in Table 9. There is a high level of heterozygosity in this species, in common with other western species, as noted above.

4.2 LL34

4.2.1 Interspecific variation

Variation among species was not assessed for this locus.

4.2.2 Intraspecific variation

4.2.2.1 *Lygus lineolaris*

A consensus reference sequence for this locus is given in Table 10.

The region contains two embedded adjacent short microsatellites, AT repeated 5 to 10 (most commonly 7) times, and GT repeated 3 to 9 (usually 6) times. Most samples were heterozygous for only one of the two repeat tracks, so the repeat length could be reconstructed by comparing forward and reverse sequences. There are few other shared polymorphisms. C is present instead of T at position 401in many samples from Prairie Provinces and one from southwestern Ontario, always associated with GT repeat length of 7. GG occurs instead of AA at position 402-403 in samples from southeastern Ontario (Ottawa and counties Prince Edward, Prescott-Russell and Stormont-Dundas-Glengarry) and is present (at least one copy) in most individuals there.

Distribution of these characters is summarized in Table 11. The few additional point changes were unique to individuals.

4.3 LL42F

A consensus reference sequence for this locus is given in Table 12. This locus is apparently protein-coding, with no stops in the reading frame beginning at position 2 of the reference sequence.

4.3.1 Interspecific variation

Base positions differing among *Lygus* species are given in Table 13. There are few concordant patterns evident. *L. abroniae* and *oregonae* share 2 changes (both G to A, positions 181 and 289) and *L. unctuosus*, *L. plagiatus*, *L. rubroclarus* and *L. borealis* share in various combinations changes at positions 110, 196, 316, 319, 331, and 420. The monophyly of *Lygus* and its separation from *Henrilygus* and *Nonlygus* is strongly supported (31 characters, data not shown).

4.3.2 Intraspecific variation

4.3.2.1 *Lygus lineolaris*

Shared intraspecific variation in *L. lineolaris* is given in Table 14. A change at position 326 results in a methionine/leucine alternation in the translated amino acid sequence. All other changes are silent with respect of amino acid translation. A 6-base insertion occurs after position 537 at eight localities in west Quebec and in Renfrew County in Ontario. No other geographic trends were noted. No linkages among changes are apparent except that all samples with C (either as homozygote or as heterozygote S) at position 104 also have G (or R) at position 185 and A (or M) at position 407.

4.3.2.1 *Lygus borealis*

The sample size for this species is not large (15 collections). No sequence variation was found among *L. borealis* individuals examined.

4.4 LL56F

Table 15 gives the majority consensus for this locus.

4.4.1 Interspecific variation

Table 16 summarizes differences among *Lygus* species. Of the observed character variation, seven positions are phylogenetically informative: position 68 partitions the species into two approximately equal groups; positions 66 and 629 both group *L. solidaginis, L. affinis* and *L. robustus*; postions 102 and 407 associate *L. abroniae* and *L. oregonae*; and A at position 357 is shared by *L. lineolaris* and *L. rubroclarus*. These groupings are consistent with relationships postulated from morphological evidence. The monophyly of *Lygus* and its separation from *Henrilygus* and *Nonlygus* is strongly supported (19 characters, data not shown).

4.4.2 Intraspecific variation

4.4.2.1 *Lygus lineolaris*

There is relatively little intraspecific variation in this locus. Table 17 summarizes character states. Most samples from southwestern Ontario have the combination CAA at positions 112, 202 and 582, while those from the Bruce Highlands and eastern Ontario have TGG. Two eastern Ontario samples near the St. Lawrence River have the pattern seen in southwestern Ontario. Prairie samples all have TG for positions 112/202 but a mix of G or A at position 582. Samples from Pennsylvania to Colorado are similar to those from eastern Ontario, except a sample from central Illinois, corresponding to the southwestern Ontario pattern. Four Ontario samples have intermediate states (TGR, CGG, or CGA).

5.0 Discussion

5.1 Strategy for locus design for microsatellites and flanking region sequence

During the course of this research project, various DNA extraction protocols have been used for the isolation of PCR and cloning quality DNA templates (data not shown). Initial microsatellite phage library construction used nuclear DNA purified via density gradient ultracentrifugation extractions with gram quantities of laboratory-reared insect material. DNA library construction requires a reasonably high quality DNA template. Restriction digests must be complete and without degraded ends. Density gradient purification of DNA yields the highest quality DNA. However, given the high equipment cost associated with ultra centrifugation techniques coupled with the difficulty in laboratory rearing insects to bulk weights of 2 to 5 g, alternate DNA purification techniques were explored. The CTAB extraction procedure has been found to be the most consistent both quantitatively and qualitatively. PCR amplifiable DNA is routinely extracted from fresh, frozen and ethanol preserved single and bulk insect specimens. The protocol has been moderately successful in recovering DNA from dried pinned individuals, although PCR target sizes are significantly reduced (data not shown). The equipment cost and time allotment for insect rearing and pre- and post-gradient processing are high, thereby negating the minor gains in library efficiency compared to the CTAB procedure.

Methods for microsatellite library construction previously used (Taylor et al., 1994, Sunnucks et al., 1996) was based upon size selection of 300 base-pair (bp) to 500 bp target fragments of genomic DNA that was simultaneously restriction digested with three blunt ended 4-base cutters (Rsa-I, Hae-III, Alu-I) and ligating into a blunt ended CIP-treated linearized plasmid vector. Colonies are then screened with CA and GA copolymers. Our initial attempts at repeating this approach yielded less than 1500 recombinants per library of which less than five colonies were found to contain microsatellite tracts. Moreover, upon sequencing these microsatellite recombinants, all were found to contain little upstream flanking sequence useful for PCR primer design. Given that 300 to 400 base pairs of upstream sequence is often needed for PCR primer design, an alternate cloning strategy capable of generating both significantly higher recombinant numbers and microsatellite clones containing longer flanking sequences was developed. Also, since microsatellite tract flanking regions have been found to contain polymorphic sites (Orti, G., et al., 1997), the flanking regions also provide sequence variation for phylogenetic analysis.

Ligation reactions involving sticky ended targets are significantly more efficient then those using blunt ended targets. The addition of adapters (Taylor et al., 1994) to blunt ended restriction

enzyme digestion products proceeded by ligation to lambda gt10 EcoRI arms still reduces cloning efficiency, since the first step ligation (EcoRI / Sma I adapter to Alu I and Hae III digestion products) involves blunt ended reactants. In our initial approach, genomic DNA was digested with Tsp 509 I, a four-base restriction enzyme that generates a cohesive end compatible with EcoRI sticky ends. This digested DNA was then ligated to lambda ZAP-II EcoRI cloning arms. Cloning efficiency was significantly increased with library titres ranging from 600,000 to 1,800,000 plaque forming units per ml with an average of 50,000 phages plated onto each of six to nine plates. Each plate was found to contain 20 to 50 positive plaques resulting in as many as 150 recombinant phage clones recovered per library. However, numerous additional positive plaques (upwards of 100) were not pulled since phage purification and subsequent phagemid excision of such large numbers of recombinants is difficult and probably unnecessary. Although no size fractionation was done for the Tsp 509 I digest, most recombinant clones processed to the pBluescript phagemid stage contained inserts of at least 700 bps in length, suggesting that the high titre of the library was not the result of an over-representation of very short and unusable cloned sequences. Upon sequencing a subset of the phagemid preparations containing inserts ranging from 400 bp to 800 bp, most were found to be truncated to one side of the microsatellite tract, often with less then 60 bp of polynucleotide tract flanking sequence and therefore useless for cross-track PCR primer design. Phagemids containing larger inserts up to 2000 bp (the maximum LI-COR autosequencing capacity) were somewhat more useful, with approximately one third of these clones having a minimum of 300bp to 500bp of tract flanking sequence. Some of these larger clones were used successfully for PCR primer design for microsatellite loci.

However, given the elevated numbers of truncated clones in the Tsp 509 I library, it was decided that a more appropriate strategy for library construction was to use the restriction enzyme EcoRI , a 6-base cutter, to digest genomic DNA in order to generate recombinant phagemid clones containing a higher proportion of large microsatellite inserts. The lambda ZAP-II cloning system has a maximum insert capacity of 10 Kb (kilobase pairs). Phagemid excision clones containing inserts ranging in size from 800 bp to 2000 bp were sequenced. About seventy percent of the remaining phage clones processed to the pBluescript phagemid excision stage contained inserts ranging from 3 Kb to 8 Kb in size. Subsequent subcloning into Litmus 28 plasmid vector of selected restriction fragments of these phagemids was very successful in generating subclones with inserts ranging from 800 bp to 2000 bp, many of which had the polynucleotide tract flanked by several hundred nucleotides. Subclones that were sequenced and found to have the polynucleotide tract localized to one end of the insert were re-subcloned using a different restriction enzyme or combination of restriction enzymes to release a new target fragment from the original large 3 Kb to 8 Kb phagemid.

Although this top-down strategy requiring various subcloning experiments is labour intensive, it provides the researcher with the flexibility to both create a microsatellite locus of a workable size, and to target the flanking regions for sequence variation. As many as thirty percent of the original phage clones from the EcoRI library have generated microsatellite loci with sufficient length of flanking sequence useful for PCR primer design. Alternate cloning strategies that immediately generate short target cloning inserts have the disadvantage that there is often insufficient flanking sequence for PCR primer design. If the researcher has "in house" access to autosequencing facilities where the cost per sequencing reaction can be reduced to a minimum (approximately $10), then microsatellite libraries can be constructed from 800bp to 2000bp gel purified EcoRI fragments. One hundred to three hundred phagemid excision clones can be

sequenced directly and those clones yielding sufficient flanking sequence used for PCR primer design. Recently, we have applied a "primer-walking" strategy to sequence the entire large (3Kb to 8 Kb) phagemid excision, avoiding subcloning.

The flanking region loci we have developed for *Lygus* using this strategy, as well as those developed for various aphid groups (unpublished data), provide information at various degrees of resolution from tribal to population levels. They thus provide a bridge between conserved loci commonly used for elucidating higher level relationships, and hypervariable regions which are useful for defining populations, but for which primers that work across taxa are difficult to design.

5.2 Interspecific patterns

The large number of morphologically very similar species in the genus *Lygus* presents problems for identification and for defining a sufficiently large set of characters useful for phylogenetic analysis. The phylogenetic reconstruction given by Schwartz and Foottit (1998) is based on several reweighting iterations of a fairly homoplasious character-state matrix containing a number of arbitrarily partitioned continuous characters. Thus, Reconstructions based on sequence data for mitochondrial large ribosomal subunit, mitochondrial cytochrome oxidase 1, nuclear small ribosomal subunit, and nuclear ribosomal internal transcribed spacer 1 (unpublished data) strongly support the monophyly of *Lygus* and its separation from *Henrilygus* and *Nonlygus*. They also support the basal placement of *L. plagiatus*, *L. unctuosus*, *L. rubroclarus*, *L. lineolaris*, and *L. vanduzeei*, clustering of all of these species as a single basal clade. Resolution of the relationships among these species and among the remaining species is rather poorly resolved, with the exception that *L. abroniae* and *L. oregonae* are closely related (in agreement with their similar habitat and host associations), as are *L. hesperus* and *L. pratensis*. The clearest point of disagreement is the placement of *L. borealis*, being basal among the species outside the *L. lineolaris* group according to the molecular data, while it is placed terminally by the morphological data set.

The microsatellite flanking region sequences treated here have provided some additional characters. Given the number of species involved, the number of informative (shared) characters is insufficient to resolve the ambiguities. However, the data does support previously determined general patterns, as well as raising some questions. Figure 2 maps several informative characters onto a tree extracted from the tree given by Schwartz and Foottit (1998). Points of agreement include:

1. The monophyly of *Lygus* and its separation from *Henrilygus* and *Nonlygus* is well supported (a total of 69 nonhomoplasious characters among the loci LL13F, LL42F and LL56F).

2. The separation of a group containing *L. plagiatus*, *L. unctuosus*, *L. rubroclarus*, *L. lineolaris*, and *L. vanduzeei*, from the remaining species is supported, although the boundaries of this separation are not unambiguously delineated; the basal position of the *plagiatus* group is supported (state of most separating characters in *plagiatus* group shared with outgroups, data not shown).

3. Several characters (2 in LL42F, 2 in LL56F) relate *L. oregonae* and *L. abroniae*.

4. Two LL56F characters suggest *L. atriflavus*, L. *solidaginis* and *L. robustus* are related.

5. One character in LL13F supports the clustering of *L. ceanothi* and *L. humeralis.*

Differences from the morphological data are the following:

1. *L. borealis* is more similar to the *lineolaris* group than the remaining species (especially evident in LL42F), as is also indicated by mitochondrial sequences.

2. *L keltoni* is not nested within a clade containing *L. atriflavus*, L. *solidaginis* and *L. robustus* (see point 4. above).

3. Three characters in LL56F (177, 180, 219) are shared by *L. rubrosignatus*, *L. keltoni*, some individuals of *L. hesperus*, and by a sample originally identified as *L. perplexus*, considered by Schwartz and Foottit (1998) as a synonym of *L. rugulipennis*; also the *L. perplexus* sample and some *L. hesperus* share C for A at position 175. The postulated synonymy of *L. rugulipennis* and *L. perplexus* should be re-examined. Also material identified as *L. hesperus* should be reassessed; more than one species may be included among these samples (note that the all of the character states for *L. hesperus* differing from the majority consensus are linked within individuals).

The extensive heterozygosity in LL13F shared by a number of species raises a number of possibilities. Some species showing extensive heterozygosity, such as *L. keltoni* and *L. hesperus* show approximately equal contribution of the two alleles, but many have unbalanced base ratios (Table 5). The apparent imbalance in the ratios of the heterozygous sites in these species suggests the possibility that this group of species may have originated from a polyploid ancestor or through duplication of a single chromosome or chromosome fragment. Given the balanced heterozygosity at other loci, the later hypothesis is more likely. Future work will involve karyotyping in order to resolve this. The state of these characters in the *plagiatus* group is identical to that in the outgroups, and so may be taken to be the primitive condition. Several samples (*L. mexicanus*, *L. rubrosignatus L. humeralis*, *L. ceanothi*, *L. rugulipennis*) which are homozygous at all or most of these positions have T replacing A at position 373 and C for G at position 388, and the latter 3 have C for T at position 448, all forms compatible with the states in the heterozygous samples. Most of these species are aggregated towards the middle of the tree base on morphology, while the heterozygotes tend toward the distal branches. This suggests that the heterozygous species may have arisen as a hybrid between a basal species and a member of the *mexicanus* group. However, we do not have data on the degree of intraspecific variation at these sites for most species, so that the observed patterns may be in part a sampling artefact. Additional sampling of several of these species is required.

5.3 Patterns within *Lygus lineolaris*

The most significant pattern with specimens identified morphologically as *Lygus lineolaris* is the strong separation in locus LL13F (14 characters) of many specimens from Ontario and the mid-western USA. This group is apparently stable: all individuals but one are homozygous at all defining sites, and all variation within groups is independent (*i.e.* sites varying within group A are largely uniform within group B and *vice versa*). In fact, the difference between the groups is as great as that between any pair of species. The virtual absence of evidence of hybridization in

combination with the presence of individuals of both groups at a given location suggests reproductive isolation. However, this strong separation is not evident in other loci. All shared point mutations in LL34 occur in both groups, as do most changes in LL42F. Some support is given by LL56F: most samples with C (or Y) at position 112 and A (or R) at position 202 are group A, while these states are rare in group B individuals. Polymorphic states at other positions in LL56F are restricted to group B individuals.

If the two groups are indeed reproductively isolated, then it would seem to be a relatively recent phenomenon, with a large number of polymorphic sites inherited from a common ancestral population. Alternatively, individuals heterozygous at locus LL13F (or some gene physically linked to it) may be lethally disadvantaged in some way.

Alleles for other polymorphic characters are distributed widely, indicating that in the long term there is extensive gene exchange throughout the range of the species. However, there are regional differences in allele frequency for some characters, suggesting that natural migration by individuals is rather limited. For example, frequency of occurrence of C at position 412 of locus LL13F ranges from 24% in south-western Ontario to 54% in west Quebec; G at position 385 of this locus is widespread in most of Ontario and west Quebec, but not found in any sample from the prairies or from extreme eastern Ontario (despite more intensive sampling in the latter region); and GG replacing AA at positions 402 and 403 of locus LL34 occurs in most individuals in eastern Ontario, but is absent elsewhere. The historical movement of populations cannot be determined from the present data. Current distributions may be the result of natural migration, but extensive mixing of populations by human transportation may have occurred. Mitochondrial genes (which are maternally inherited, and thus not subject to reshuffling by sexual recombination) may provide a better understanding of these historical factors.

In general, prairie populations are more uniform than those in Ontario and west Quebec. Of particular note is the greater degree of heterogeneity among samples from Renfrew County in Ontario and the adjacent Pontiac region of Quebec, as compared to any other area of comparable size. Agriculture is less extensive in these regions, and most of the collections were from non-agricultural habitats. Collections from non-crop settings in other areas were usually in close proximity to cultivated fields. Future collection in other non-agricultural areas is required to determine if genetic diversity in these areas is greater than in agricultural areas in general. If this is so, it remains to be determined whether this is the result of selection for particular crop-adapted genotypes, or the result of artificial dispersal of genotypes present in agricultural areas by historical accident.

6.0 References

Goldstein, D. B. and C. Schlotterer. (Eds.). 1999. *Microsatellites. Evolution and Applications.* Oxford Univ. Press. 352 pp.

Orti, G., D. E. Pearse and J. C. Avise. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proc. Natl. Acad. Sci. USA.* 94: 10745-10749.

Queller, D. C., J. E. Strassmann and C. R. Hughes. 1993. Microsatellites and kinship. *Trends. Ecol. Evol.* 8: 285-288.

Roderick, G. K. 1996. Geographic structure of insect populations: Gene flow, phylogeography, and their uses. *Ann. Rev. Entomol.* 41: 325-352.

Schwartz, M. D. and R. G. Foottit. 1998. Revision of the Nearctic species of the genus *Lygus* Hahn, with a review of the Palearctic species (Heteroptera: Miridae). *Memoirs on Entomology, International. Vol. 10.* Associated Publishers. 428 pp.

Sunnucks, P., P. England, A. C. Taylor and D. F. Hales. 1996. Microsatellite and chromosome evolution of parthenogenetic *Sitobion* aphids in Australia. *Genetics* 144: 747-756.

Tautz, D. 1989. Hypervariablity of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research.* 17: 6463-6471.

Taylor, A. C., W. B. Sherwin and R. K. Wayne. 1994. Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat *Lasiorhinus krefftii*. *Molecular Ecology* 3: 277-290.

Locus	Primers	Primer sequence	PCR conditions (x 35 cycles)	Product size
LL13	LL13F-3F	CAA GAA CCT GGG TGT CAT AAT AAC	95°C/60s + 52°C/60s + 72°C/90s	700 bp
	LL13F-3R	CGT ACA CGC TCT CCA ACC ATA C		
LL34	LL34F-1F	TCT CAA CGC ACC TCA ACT TCA	95°C/60s + 53.5°C/60s + 72°C/120s	1100 bp
	LL34R-2R	CAA AAT TTG ACC CCT TGG AGT AAC		
LL42	LL42R-3F	CGC CGA GAA GTT TCC TGT TGT A	95°C/60s + 52°C/60s + 72°C/90s	550 bp
	LL42F-6R	CGC CCT TCA ATT GTT CAT CG		
LL56	LL56F-2F	GTG GGG GTT AAA CTT ACG AAA TC	95°C/60s + 52°C/60s + 72°C/120s	780 bp
	LL56F-2R	GAT CCC AGA GGT AGT AGA CAT TCC		

Table 1. PCR primer sequences and amplification conditions. All reactions preceded by a denaturation step of 95°C for 4 minutes, and followed by a final extension at 72°C for 8 minutes.

Locus	Primers	Primer sequence	Product size
LL13	LL13F-4F	CGT CAA GTG GCG GAA ATA TG	500 bp
	LL13F-4R	CCA ATA ACG ATG TTT TAA GCG AG	
	LL13F-5R	CGG AAA GCT GTA ACA CTT GTG C	
LL34	LL34R-7F	GTG TGT GTG GGT CAT TTC AGA	320 bp
	LL34R-4R	CGC ATT TCA CAT ATT TCA CT	

Table 2. Internal sequencing primer sequences. LL13F-5R used with LL13F-4F for cross species comparisons, and LL13F-4R used with LL13F-4F for within-species comparisons in *L. lineolaris*, *L. borealis* and *L. keltoni*. Sequencing primers for LL42 and LL56 are same as PCR primers.

```
TGACACGTTA GACTGGAATC GTCAAGTGGC GGAAATATGT AAGAAGGTGT ATGGTAGCCT  60
GTATCAGTTG CGTAGAATTG CCTTCGATTT TCCAAAGCAT GTCAGGACGC AATTGGCTCA 120
GGCCCTCCTG GTGCCTTACT TTGAATATGC GCCTTTGGCT TTCTGCGATT TGAATAACGA 180
GCAAATGGGC AGACTACAAA AAACGCTCAA CTGTGTTGTG CGTTTTGTAT GTCGTCTAAG 240
GCTGGATGCG CATGTTACGC CGGCGTACTT GGAGCTTGGG TGGCTTAAAA TGGAGGAGAG 300
GAAAAGGTTG GCGGTTGGAG CAATGTTGTT CAAAATTCTC AAATTCAGGA AACCTCAGTA 360
CTTATATAAC CAATTTAGGT ATCTTTCGTC TGTGCATACG GTTTCAACGC GCAAAGCGGC 420
TACTACACTC CAGATACCGA AGCACAATAC GGTTTTATTC AGCAGATCGT TCATAATGCA 480
AGCCATTGAA ATGTATAACT CTAATGCAGA GATTTTTGAT TTAAGCACAA GTGTTACAGC 540
TTTCCGAAAC TCGCTTAAAA CATCGTTATT GGAGAGGTAT ATGTAGTAGG ATGAGTTGAT 600
AAGTTAAACC CACAAGTGCA GCGTTTTCCG ACCATGTACT GCTGCTCGCT GTTATATAAT 660
TGCGGGCGGT GGT                                                     673
```

Table 3. Reference majority consensus sequence for LL13F based states found in *Lygus lineolaris*.

Species	State value at given postion
	1 4 55 56 65 97 115 177 188 201 260 289 306 314 320 322 328 346 352 369 373 388 391 397 400 430 439 445 448 451 457 459 473 476 506 526 544 551 584 602 609 618 646 654 656 658 660 662 663 667 669 671 678
abroniae (1)	??CACRGRRACRRGGRRYARWSYYRCRYYKGWAT????????????????????
oregonae (1)	??CACRGRRACRRGGRRYARWSYYRCRYYKGWAT????????????????????
striatus (1)	??CACRGRRACRRGGRRYARWSYYRCRYYKGWAT????????????????????
robustus (1)	??CACRGRRACRRGGRRTARWSYYRCRYYKGWAT????????????????????
solidaginis (1)	??CACRGRRACRRGGRRYARWSYYRCRYYKGWAT????????????????????
atriflavus (2)	??CACGGR$^{G}_{R}$ACRGGGRRYARTSY$^{Y}_{T}$RCRYYKGWAT????????????????????
potentillae (1)	??CACRGRGACRGGGRRYARTSYYRCGYYKGWAT????????????????????
shulli (1)	TCCA?GGAGACAGGGRAYARAGYYACRTTTGWATCCYTTTAGTGTTTATGGA?
keltoni (14)	cCCAC$^{A}_{R}$GR$^{R}_{G}$ACrGGGga$^{Y}_{C}$A$^{R}_{G}$A$^{S}_{G}$CC$^{R}_{A}$CatTTG$^{W}_{A}$ATCC$^{Y}_{T}$T$^{C}_{\underline{Y}}$T$^{C}_{\underline{T}}$AGT$^{A}_{\underline{G}}$a$^{G}_{\underline{K}}$at$^{A}_{W}$G$^{C}_{\underline{S}}$$^{A}_{G}$$^{G}_{T}$
hesperus (2)	?CCAG$^{G}_{A}$GA$^{G}_{R}$AC$^{A}_{G}$GGGA$^{R}_{A}$YARAGY$^{T}_{Y}$$^{R}_{K}CA^{C}_{T}TTG^{W}_{A}$ATCCCTTTAGTATTTATGGGT
elisus A (1)	??CACGGAGACAGGGAGCAAWSTTGCGCTGGTAT????????????????????
elisus B (1)	TCCACGGAGACAKGGGRCAAAGTTACGYTTGTAT?C??TTAGTRTTTAT????
plagiatus (1)	??CACGGAGACAGGGAGCAAAGTTGGGCTGGTAT????????????????????
rubroclarus (1)	??CACGGAGACAGGGAGCAAAGTTGCGCTGGTAT????????????????????
unctuosus (1)	??CACGGAGACAGGGAGCAAAGTTGCGCTGRTAT????????????????????
adspersus (1)	??CACGGAGACAGGGAGCAAAGTTGCGYTGATAA????????????????????
pratensis (1)	??CACGGAGACAGGGAGCAAAGTCACGTTGGTAT????????????????????
borealis (10)	TTCTCGAAGACAGaGAGCAAAGTTGCGCTGGTATGCCTTTGGGG$^{T}_{G}$TTATGGAT
vanduzeei (4)	TTCA$^{G}_{C}$GGAGACAGGGA$^{G}_{R}$CAAAGTT$^{R}_{G}$CGCTGGTATGCCTTTAGTGTTTATGGAT
lineolaris B (385)	TC$^{T}_{C}$A$^{G}_{C}$GGAGACAGGGAGCAAAGTTGCGCTGATAAGCCTT$^{A}_{\underline{T}}$A$^{A}_{\underline{G}}$TGTTtaTGGAT
lineolaris A (38)	TCCACGGAGG$^{G}_{\underline{S}}$AGGGAGTGATCTTGAGCTGGTGTGTCG$^{T}_{\underline{A}}$TAGTGTTTGAGGAT
rubrosignatus	??CACGGAGACAGGGAGTAATCTTGCGCTGGTGT????????????????????
mexicanus (1)	??CACGGARACAGGGAGTAATCTTGCGCTGGTAT????????????????????
humeralis (1)	??CACGGAGACAGGGAGTAATCTTGCGCCGGTAT????????????????????
ceanothi (1)	??CACGGAGACAGGGAGTAATCTTGCGCCGGTAT????????????????????
rugulipennis (1)	??CACGGAGACAGGGAGTAATCTTGCGCCGGTAT????????????????????

Table 4. Interspecific variation in LL13F. Matrix of states corresponding to positions 1, 4, 55, 56, 65, 97, 115, 177, 188, 201, 260, 289, 306, 314, 322, 328, 346, 352, 369, 373, 388, 391, 397, 400, 412, 430, 439, 445, 448, 451, 457, 459, 473, 476, 506, 526, 544, 551, 599, 606, 615, 643, 651, 653, 655, 657, 659, 660, 664, 666, 668, and 673 of reference sequence (see Table 3). Number of individuals sequenced for each species indicated in parentheses following name. Uppercase ambiguity codes represent heterozygosity within sampled individuals (W=A/T, S=C/G, R=A/G, K=G/T, Y=C/T, M=C/A). For sites showing variation among individuals within a species or group, if less than 20% of individuals sampled show variance from the majority, the majority form is shown in lower case; if greater than 20% of individuals vary from majority, the two most common (or only) forms are shown in small font, the majority above; underlined codes indicate that a small proportion of individuals showing a third compatible state are present (*e.g.* $^{T}_{\underline{Y}}$ means at least 1 individual is homozygous for C at this site, $^{T}_{C}$ means most individuals are homozygous but at least 1 individual is heterozygous at this site [Y]). (Within-group variation is detailed in Tables 6 through 9.)

position	97	177	188	289	306	322	328	346	369	373	388	391	397	400	439	445	448	451	459
base 1	G	A	G	A	G	A	G	T	A	T	C	T	T	G	G	C	C	G	T
base 2	A	G	A	G	A	G	A	C	G	A	G	C	C	A	A	T	T	T	A
abroniae	5	5	6	4	5	4	4	4	6	4	5	4	5	5	5	5	4	5	2
oregonae	2	3	3	1	2	1	2	2	2	2	3	2	2	2	2	2	2	2	2
striatus A	4	7	5	4	4	5	5	5	5	5	4	5	6	5	5	5	2	5	4
striatus B	3	5	5	2	3	3	3	4	3	4	4	4	5	4	4	4	4	5	3
robustus A	6	8	8	8	8	8	7	7	7	8	6	8	8	8	7	7	7	9	-
robustus B	3	2	2	2	5	2	2	8	3	2	4	2	2	2	3	3	3	2	2
potentillae	6	7	8	5	x	5	8	8	5	8	7	7	8	8	6	6	5	8	8
elisus A	10	10	10	10	3	10	10	2	10	3	4	10	10	10	10	10	0	10	10
elisus B	10	10	10	10	x	0	x	0	10	0	0	10	10	0	10	x	0	0	10
rugulipennis	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	3	10	10
shulli	10	10	10	10	10	x	0	x	x	0	0	x	x	10	x	0	0	0	x
hesperus	x	10	10	x	10	x	x	x	x	0	0	x	x	x	0	x	0	0	x
keltoni	x	x	x	x	10	0+	0+	x	x	0	x	0	0	x	0+	x	0	0	x

Table 5. Approximate proportions of components for LL13F at positions heterozygous in several species. Value in body of table indicates 10 times approximate relative proportion of base 1 to total peak height for that position (e.g. for *L. oregonae* at position 97, value of 2 indicates apparent mix is 20% G and 80% A) (signal corrected for relative lane intensity). 'x' represents heterozygous position for which relative intensity has not been determined. The component selected as 'base 1' is that matching the reference sequence. '*L.robustus* B' is a sample originally determined as synonym '*L. columbiensis*', '*L. striatus* B' is a sample originally determined as synonym '*L. nigropallidus*'.

state values	Sample number (Location)
7 37 52 65 90 109 147 152 179 250 281 289 290 338 388 404 405 410 430 437 512 541 551 557 569 600 602 607 632	
GGTCTGTCGGTAACCTCCACATGATTAAC	99em60 (N. Woodslee ON), 99em62 (Dresden ON), 99em150 (Waco NE)
G**A**TCTGTCGGTAACCTCCACATGATTAAC	99em53 (Rushton Cnrs ON), 99em65 (Harrietsville ON), 99em66 (Harrietsville ON), 99em81 (Joyceville ON)
G**R**TCTGTCGGTAACCTCCACATGATTAAC	99em30 (Price Cons.A. ON), 99em750 (Glen Robertson ON), 99em753 (Carlsbad Springs ON)
??TCTGTCGGTAACCTCCACATGATTAAC	01em22.1 (Wellandport ON), 01em32.1 (Melbourne ON), 01em77.2 (Lombardy ON), 01em83.2 (Bloomfield ON), 01em86.2 (Price Cons.A ON), 01em 87.1 (Kaladar ON)
GGTCTGTCGGTA**T**CCTCCACATGATTAAC	99em67 (Delhi ON)
GGTC**Y**GTCGGTAACCTCCACATGATTAAC	99em80 (Ernestown ON)
G**R**TCTGTC**K**GTAACCTCCACATGATTAAC	99em141 (Farmer City IL)
G**R**TCTGTC**KS**TAACCTCCACATGATTAAC	99em746 (Grantley ON)
??TCTGTCG**S**TAACCTC**T**ACA**C**GATTAAC	01em80.2 (Colebrook ON)
??TCTGTCG**S**TAACCTCCACATGATTAAC	01em82.2 (Roblin Mills ON)
G**R**TCTGTCG**C**TAA**T**CTCCACATG**G**TTAA**T**	99em52 (London ON)
G**R**TCTGTCG**S**TAACCTCC**C**CATGATTAAC	99em126 (Newcastle IN)
GGTCTGTCG**S**TAACCTCC**C**CATGATTAAC	01em81.2 (Camden East ON)
??TCTGTCGGTAACCTCC**C**CATGATTAAC	01em11.1 (Warkworth ON)
??TCTGTCGGTAACCTCC**C**C**GT**GA**C**TAAC	01em83.1 (Bloomfield ON)
G**AY**CTGTCGGTAACCTCC**M**CATGATTAAC	99em54 (Wheatley ON)
G**A**TCTGTCGGTAACCTCCACATGATTAAC	01em 118.3 (Long Sault ON)
G**A**TCTGTCGGTAACCTCCACATGAT**WWC**C	99em57 (Cedarhurst ON)
SATCTGTCGGTAACCTCCACATGATTAAC	99em64 (Warwick ON)
G**A**TCT**SYMGS**TAACCT**Y**CACATGATTAA**S**	99em77 (Northumberland Co ON)
G**A**TCTGTCGGTAACCTCCACAT**K**ATTAAC	99em146 (Bettendorf IA)
G**A**T**G**TGTCGGTAACCTCCACAT**T**ATTAAC	00em11a.1 (Calabogie ON)
??TCTGTCGGTAACCTCCA**S**ATGATTAAC	01em11.2 (Warkworth ON)
??TCTGTCGG**AT**ACCTCCACATGATTAAC	01em28.2 (Harrietsville ON)
??TCTGTCGGTAAC**S**TCCACATGATTAAC	01em84.1 (Melville ON)
GGTCTGTCGGTAAC**SC**CCACATGATTAAC	01em84.3 (Melville ON)

Table 6. Variation within *L. lineolaris* LL13F group A. States corresponding to 7, 37, 52, 65, 90, 109, 147, 152, 179, 250, 290, 338, 405, 430, 551, 557, 600, 602, 607, 632 of reference sequence (see Table 3).

	Character states at position	Canadian Prairies	Southwestern Ontario	southeastern Ontario	West Quebec	Quebec (Eastern Townships)	US w of Miss. R.	US e of Miss.R.
		n = 57	n = 70	n = 157	n = 66	n = 9	n = 5	n = 6
49	G:T	0.787:0.213	0.794:0.206	0.849:0.151	0.869:0.131	0.813:0.187	1:0	0.833:0.167
55	T:C	0.802:0.198	0.806:0.194	0.790:0.210	0.795:0.205	0.938:0.062	1:0	0.833:0.167
65	C:G:A:T	0.926:0.065 :0:0.009	0.984:0.005 :0:0.005	0.936:0.054 :0.007:0.003	0.914:0.052 :0.017:0.017	1:0:0:0	1:0:0:0	1:0:0:0
385	T:G	0.988:0.012	0.878:0.122	0.922:0.078	0.846:0154	1:0	1:0	1:0
412	T:C	0.798:0.212	0.763:0.237	0.638:0.362	0.458:0.542	0.625:0.375	1:0	0.6 : 0.4
415	A:T	0.991:0.009	0.950:0.050	0.952:0.048	0.992:0.008	1:0	1:0	1:0
581	No ins : insert	0.777:0.223	0.825:0.175	0.792:0.208	0.820:0.180	0.875:0.125	1:0	0.786:0.214
599	A:T*	0.833:0.167	0.727:0.273	0.779:0.221	0.774:0.226	1:0	0.75:0.25	0.667:0.333
615	A:G*	0.929:0.071	0.923:0.077	0.912:0.088	0.972:0.028	1:0	1:0	1:0

Table 7. Variation within *Lygus lineolaris* LL13F group B. Frequency (by geographic region) of state over sampled haplotypes, assuming diploid genome with single copy of locus in haploid chromosome set. *Values for positions 599 and 615 are based on individuals homozygous for insertion/deletion at position 581 (unidirectional sequencing across indel in heterozygous individuals yields unreadable superimposed sequence beyond indel site).

States at variant sites	Sample number
TCAGT	99em41, 95em219, 99em696, 2003LB271, 2003LB348
TC**RR**T	99em686
T**T**AGT	95em435.3
CCAG**G**	95em435.4

Table 8. Variation in LL13F among *L. borealis* samples. States corresponding to positions 1, 58, 314, 636, and 653 of reference sequence.

AGRGAYRGRAYWTSCAATTTAGGT	95em555b
RRRRAYRSRATWYYTGAKAKASAG	95em560a.1
RGRGAYRSRATWTYTGAGATACAG	95em560a.4
AGAGACRGRATWYCCRAGATWCGT	95em562.5
ARRGACRSAATWTYYAAKATWSGG	pgm97-2.1
ARRGACGSAATWYYYAAKATAGGG	pgm97-3.1
ARRGACRGAATATCCAAGATWCGG	pgm97-4.2
AGRGAYGSAATWYCCAAGWTACAG	pgm97-4.3
AGRGACRSAATATCTAAGATACAG	pgm97-5.1
AGRGAYRSAATWTCTAAKATACAG	pgm97-5.2
RRRRRYGSRRYAYYYAAGATACGT	pgm97-6.1
ARRGAYRSAATWYCCAATATWCGG	pgm97-6.2
RGRGAYRSAATWCCTATTT??GGT	99em152

Table 9. Variation within *L. keltoni* for locus LL13F. States corresponding to positions 97, 188, 289, 322, 328, 346, 369, 388, 400, 439, 445, 459, 544, 584, 599, 651, 654, 655, 657, 659, 660, 664, 668, and 674 of reference sequence (see Table 2).

```
GCACCTCAAC TTCAAGTTTT AGCGAATTTT CGAAGAAAAT TAAGCTATTA CAATGGCTTC  60
AAAATTTTTG TAATGCAAAT TTTTGAAAAT TTTCTTTATG GCAATCCGTT TCGGACCATT 120
CTAAAAGGAT TCCTGTAATT TAACATTTTG TCATATATAT ATATATTTGT GTGTGTGTGT 180
GGGTCATTTC AGACGATATC TGCCCGAGGG ATCGAGATAA AAAATCCAAA GAAAATTATA 240
AAATGTTAAT AATATCAAAA GAAAGGATGA GTTCGT                           276
```

Table 10. Majority consensus sequence for locus LL34F in *Lygus lineolaris*. Length-varying

dinucleotide repeats are highlighted in bold face.

Collection #	Region	AT repeat length	GT repeat length	401	402-403
2003em14	se Ontario	7	3/5	T	RR
99em55	sw Ontario	9	5/6	T	AA
2003em9	se Ontario	6/10	6	T	GG
2003LB190.2	Canadian Prairies	6/8	6	?	?
99em51.1	sw Ontario	6/7	6	T	AA
99em25	se Ontario	6/7	6	T	GG
2003LB410	Canadian Prairies	6/7	6	T	AA
2003LB330	Canadian Prairies	6/7	6	?	?
99em121	eastern US	7	6	T	AA
99em141	eastern US	7	6	T	AA
99em691	Canadian Prairies	7	6	T	AA
2003em30	west Quebec	7	6	T	AA
99em738	ne Ontario	7	6	T	AA
2003em13	se Ontario	7	6	T	RR
2003em15	se Ontario	7	6	T	RR
2003em22	se Ontario	7	6	T	GG
2003LB190.3	Canadian Prairies	7	6	?	?
2003LB210.2	Canadian Prairies	7	6	T	AA
2003LB222.2	Canadian Prairies	7	6	T	AA
2003LB292.1	Canadian Prairies	7	6	?	?
2003LB335.1	Canadian Prairies	7	6	T	AA
2003em33	west Quebec	7	6	T	AA
2003em17	se Ontario	7	6	?	?
2003em11	se Ontario	7	6/8	T	AA
2003LB222.1	Canadian Prairies	7	6/9	?	?
2003LB222.3	Canadian Prairies	7/8	6	?	?
2003LB292.3	Canadian Prairies	8	6	T	AA
2003em4	se Ontario	8	6	T	GG
99em717	northern Ontario	8/10(C)*	6/7	Y	AA
99EM51.2	sw Ontario	4	7	C	AA
2003LB302.2	Canadian Prairies	7	7	Y	AA
99em689	Canadian Prairies	7	7	Y	AA
99EM73	sw Ontario	7	7	T	AA
2003EM28	west Quebec	7	7	T	AA
2003EM23	se Ontario	7	7/8	T	GG
99EM701	Canadian Prairies	7	7/8	Y	AA
2003EM5	se Ontario	7/8	7	T	AA
99EM697	Canadian Prairies	9	7	C	AA
2003LB206	Canadian Prairies	8	8	?	?
2003EM31	west Quebec	9/11	6	T	AA

Table 11. Shared variation in LL34 for *Lygus lineolaris*. *Penultimate AT repeat unit for sample 99EM717 has C replacing T.

```
TTCATCGTGC ATGAAGGAAA GGAGATTGTC CCGTGTCGTA ACGTGCCACC TCTGAAGCCG  60
ATAAGCAAGT TGAACTACTG AATGTTATTG AGGAGATAAG GGAGGAAATT AGGAGTTCTA 120
ATCGGCAACT TCAAGAGGAG ATTGAACGCC AGGCCGGTGA GATTAATGAT CTTAAAGTCC 180
AGCTCAGTAC TTACTCAGAT TATATAGAAT CTAACAAACA ATCGTTGGAT CGTGTGGACT 240
CTTCTTTGAA GGCTCTCGCC GATAAAGTGG ACAATGTGAT GGACTGTCAG AAGGGATATG 300
ATAAGAAGCT CGAAGAGCTG ACTGAAATGA TAAACAACGT TGATCAACAA GCTCGTGAGT 360
CGTCAGTGGA AATTACTGGA TATCCAGAAA CGGAAAATGA GAATGTTCTT GAGATCGTAA 420
GAAAGATCGG CGATGCAGTA AAGTTTCCTA TTTCTGAGCA AATGCTGGAT GACTGCTATC 480
GAATTAAACC CAGGAATCCA CGACCAGGTT TCCCGGGCAC AATTGTTGTC TCTTTTGTGC 540
GAAAATTGGA TAAAAAAGGT TTTTACGCTG CTGCCTGGAA AATTAAAGAT TTCAATACTC 600
GGAACGTGGG ATTCATGCTT GGAGAATCTA ATAGAATCTA CGTTAACAAC AGTTTGACCC 660
```

Table 12. Majority consensus sequence for locus LL42F across *Lygus* species.

Lygus sp.	26	28	40	55	56	74	77	110	118	131	181	196	241	259	266	289	292	314	316	317	319	322	331	337	355	364	366	385	403	406	416	418	420	432	484	493	499	501	502
lineolaris	T	G	A	A	G	C	C	A	T	C	G	A	A/T/W	C	G	G	G	G	G	C	G	T	A	C	T	A	T	A	T	T	G	A	G	A	T	G	A	G	A
borealis	·	·	C	·	·	T	·	·	Y/C	M	A	·	T	·	·	·	·	·	T	·	A	·	T	T	·	·	·	·	·	·	·	·	·	·	·	·	G	·	T
unctuosus	·	·	C	·	·	·	·	C	·	·	·	C	T	·	·	·	·	·	T	·	A	·	T	T	·	·	·	·	·	·	·	·	A	·	·	A	·	·	·
plagiatus	·	·	C	·	·	·	·	M	·	·	·	M	T	M	·	·	·	·	?	·	A	·	W	Y	·	·	·	·	·	·	·	·	R	·	·	·	·	·	·
rubroclarus	·	·	C	·	·	·	·	C	·	·	·	·	T	·	·	·	·	·	·	·	·	·	T	T	·	·	·	·	·	K	·	·	·	·	C	·	·	·	·
vanduzeei	·	·	C	·	·	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
rufidorsus	·	·	·	·	T	·	·	·	·	·	·	·	T	·	·	·	A	·	·	·	·	·	·	T	W	T	·	·	·	·	·	·	·	·	·	·	·	·	·
humeralis	·	·	C	·	·	·	·	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
oregonae	·	·	C	W	·	·	·	·	·	·	R	·	T	·	·	A	·	S	·	·	·	G	·	T	·	·	·	·	C	·	·	·	·	·	·	·	·	·	·
abroniae	·	·	C	·	·	·	·	·	·	·	A	·	T	·	·	R	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
elisus	·	·	C	·	·	·	C/T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T/Y	·	·	·	·	·	·	A	·	·	·	·	·	·	·	·
robustus	·	·	C	·	·	·	Y/T	·	·	·	·	·	T	C/Y	K	·	·	·	R	A/M	·	·	·	T	·	·	Y	·	·	·	·	·	·	R	·	·	·	·	·
ceanothi	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
convexicollis	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
pratensis	C	T	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	·	·	·	·	T	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·
shulli	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
striatus	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
rugulipennis	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	C	·	·
hesperus	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
rubrosignatus	·	·	C	·	·	·	T	·	·	·	·	·	C	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
potentillae	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
solidaginis	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
atriflavus	·	·	C	·	·	·	T	·	·	·	·	·	T	·	·	·	·	·	·	A	·	·	·	T	·	·	·	·	·	·	·	R	·	·	·	·	·	·	·

Table 13. Variation in LL42F among *Lygus* species. Sites with variant forms unique to *Lygus lineolaris* are omitted. Positions given relative to reference sequence in Table 13. Minority base variants highlighted in bold. Dot indicates identity with *L. lineolaris*. Ambiguity codes represent heterozygous individuals, not genotypic polymorphisms (which are indicated by multiple entries).

Sample number	104	185	242	326	360	370	393	407	442	519
99em751, 2003em19, 2003em29, 2003em35, 2003em36, 2003em79, 2003em82	G	A	T	A	C	A	A	C	G	C
99em732	.	A	.	.	.	.	.	.	A	.
2003em13, 2003em72	.	R	.	.	.	.	.	.	A	.
99em743, 99em746, 99em747, 99em750, 2003em10	.	.	.	.	.	.	.	.	R	.
95em275.1, 99em734,	.	R	.	.	.	.	.	M	.	.
99em123, 99em757	S	R	.	.	.	.	.	M	.	.
99em748	S	R	W	M			.	M	.	?
95em254a.1	S	R	W	M	.	.	.	M	.	.
99em692	S	G	.	.	.	.	.		.	.
2003em11, 2003em31	C	.	.	.	.	.	.	A	.	A
99em146	.	.	.	.	M	.	.	.	R	.
99em749	.	R	.	M	.	.	.	.	.	?
95em322a.2	.	R	.	.	A	.	.	.	.	.
95em323.2, 99em749	.	.	.	M	M	.	.	.	.	.
99em745	.	.	.	M	M	.	.	.	R	.
95em326a.1, 99em126, 99em717, 99em738, 99em752, 2003em15, 2003em38, 2003em39, 2003em43	.	.	.	M	.	.	.	.	.	.
95em296a.1	.	.	W	M	M	R	.	.	.	.
2003em18, 2003em85a	.	.	W	M	.	.	.	.	.	.
95em326a.1	.	R	?	M	M	.	.	M	.	.
95em284.1	.	.	?	M	.	.	R	.	.	.
95em324a.1	.	R	?	M	.	R	R	.	.	.
95em285a.1	.	.	W	M	A	R	G	.	.	.
95em297.1	.	.	?	C	M	R	G	.	.	.
95EM284.1, 95em303a.1	.	R	?	M	M	.	G	.	.	.
95em281a.1, 95em283a.2, 95em298a.1	.	.	?	C	M	.	G	.	.	.
95em306a.1	.	.	W	C	A	.	G	.	.	.
95em347a.1	.	R	W	C	M	.	G	.	.	.
99em696, 99em735, 99em753, 2003em14, 2003em26, 2003em81	.	.	W	C	.	.	.	.	.	.
95em249.1, 99em691, 99em701, 2003em20, 2003em30, 2003em28, 2003LB190a	.	.	A	C	.	.	.	.	.	.
95em346.1	.	R	.	C	M	.	.	M	?	?
95em276a.1, 2003LB190c, 2003em21, 2003em27, 2003em34, 2003em37, 2003em41	.	.	.	C	.	.	.	.	.	.
2003em61, 2003em76	.	.	.	C	.	?	?	.	.	.

Table 14. Shared variation within *Lygus lineolaris* in LL42. Dot indicates identity with first row.

```
ACTTACGAAA TCGAGATCAT CAAAAGAATT GATGATTATT TAGTAGCAGA AAAGAGCATT  60
AGGTACGACT AATGATTTCC AGATGGAACC TGTTGATTGA AGAAAATGAC ATGCAAATAT 120
GAGCACTTAC GAGCTGATGA CAATGAACAC ATACCTATAC TTGGCATGGT ATATACCACA 180
CCCAGAGATA AGTTCATGTT GGGGAAAACT TGCGCGCCGC CCAAGTGTTT ATGAACCATG 240
GTGTTGAATA TCATAAGGTG AATGATGGAA AATGAGGGGT CCCAAGCTGC GAGGCACCAG 300
CGGTACCCGC GGTTTCCTCA TTTATTGCAA AAGGACGTCA GGATAGATAA GACACCGCTT 360
CTATTTAGGT ATAGGATGCG GGTGGGCCAT AAATTAGATG CCTTGGAAGG TTGTATATAG 420
GGCAGAAACC ACGAGCTGTA TGAGAGGGAG GTCGGTGTGA AGTCATCTAC CATTATCCAC 480
TCGAGCTCCT TTTACCCGCC ACGATCGCTC GTAATGAACG GATTACAGCA TAACAGTGGT 540
GCATGGCGGA TCTATTATGC ATGCCTTCAT TAATTATTTA GGGCATAATG AGCCTAGTGA 600
AGCCTTAACA GTCTCGCCCG TATTTGCATC ACTATGCAGG GAGATTTCCA ATGAGGGAGC 660
CGTAGTTAAG TATGGCTTGT CTCCGCTTTC CCACTCCCGA GCCTTCTACT CCGATTCTCA 720
TGCCACCCAA CGCAGGATTG GGTGATAAGC CTAATCAACC                       760
```

Table 15. Majority consensus sequence for locus LL56F for *Lygus lineolaris*.

	68	102	110	120	121-126	122	123	126>	129	133	135	136	140	158	159	159>	167	170>	175	177	180	209	219	275	302	320	354	363	370	404	437	448	449	476	516	547	582	633	645	655	676	688	695	702	707	730	733	752	762
lineolaris	A	G	T/C	T	+	A	G	–	A	G	T	G	A	T	A	–	C	–	A	C	T	A	G	–	C	A	A	T	T	T	A	G	A	A	G	C	G/A	C	T	G	C	T	T	G	T	C	A	G	C
rubroclarus	.	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
borealis	.	.	C	C	.	.	.	–	.	.	.	.	.	.	T	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
unctuosus	.	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	T	–	.	.	.	.	.	–	.	G	G	.	.	.	.	A	G	.	.	.	A	.	.	.	A	.	.	.	.	.	A	.
plagiatus	.	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
rufidorsus	.	.	C	C	–	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.
elisus	.	.	T	C	.	.	.	–	.	.	.	.	.	C	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	G	.	C	.	.
humeralis	.	.	C	C	.	T	.	A	.	A	C	.	.	C	.	–	.	*	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	C	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
hesperus	.	.	C	C	.	.	.	–	.	.	.	.	.	.	G/A	–	.	–	C/A	C/T	C/A	.	+/-	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
keltoni	.	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	T	C	.	–	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
rubrosignatus	M	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	T	C	.	–	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
rugulipennis B	C	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	C	T	C	.	–	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
rugulipennis A	C	.	C	C	.	.	.	–	.	.	.	.	.	.	G	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
abroniae	C	A	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	C	.	.	.	.	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.
oregonae	C	A	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	A	C	T	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	A
bradleyi	C	.	C	C	.	.	T	–	.	.	.	.	.	.	.	TA	.	–	.	.	.	.	.	A	T	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.
striatus	C	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	G	C	A	.	.	.	.	.	.	.	.	.
potentillae	C	.	C	C	.	.	.	–	.	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	C	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	.	.	.	.	.	.
solidaginis	C	.	C	C	.	.	.	–	G	.	.	C	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.
atriflavus	C	.	C	C	.	.	.	–	G	.	.	.	.	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.
robustus A	C	.	C	C	.	.	.	–	G	.	.	.	.	.	.	–	.	–	.	.	.	T	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	T	.	.	.	.	.
robustus B	C	.	C	C	.	.	.	–	G	.	.	.	T	.	.	–	.	–	.	.	.	.	.	–	.	.	G	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.

Table 16. Variation among *Lygus* species in LL56F. Positions relative to reference sequence given in table 15. Dot indicates identity with state in *L. lineolaris*. '*' = ACAT insertion. '>' with position number indicates insertion following given position. '*L. rugulipennis* B' originally identified as junior synonym *L. perplexus*. '*L. robustus* B' originally identified as junior synonym *L. columbiensis*.

Region: [sample]	96	112	115	119	120	150	151	152	155	159 >	202	203	173-177	232	286	403	407	453	484	560	582
wQC: [LL10, LL12, LL15, LL16, LL18]. AB: [LL23, 95EM283, 95EM284, 95EM303, 95EM324, 95EM346]. MB: [95EM420], SK: [95EM433, 95EM478. nON: [99EM732]. eON: [[99EM34, 99EM35, 99EM36, 99EM45, 99EM48, 99EM71, 99EM738, 99EM749, 99EM751, 99EM752, 2003EM49, 2003EM55] swON: [99EM28, 99EM37, 99EM38, 99EM79, 99EM84]. IN: [99EM127] IA: [99EM149] NE: [99EM151, 99EM152].	A	T	A	A	C	C	A	T	C	-	G	G	*	T	G	T	A	C	A	C	G
swON: [99EM44]	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	.	.	.	.
swON: [99EM49]	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	.	.	Y	.	Y	.
MB: [95em402, SK: [95em473]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.	Y	.
swON : [99EM75]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.
wQC : [LL20]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	T	.
wQC: [LL17]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R	.	.	Y	.	T	.
MB: [95em414]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
BC: [LL21]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.
AB: [95em285, 95EM297]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	.
eON: [LL24]	.	.	.	.	.	.	R	.	.	.	.	.	.	.	.	.	.	.	.	Y	.
AB: [95em347a]	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SK: [LL31]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.
MB: [95em429]	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
OH: [99EM125]	.	.	.	.	.	.	.	.	.	.	.	.	**	.	.	.	.	.	.	.	.
2003EM46	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
swON: [99EM50]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
eON: [99EM77, 2003EM43]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	R
2003EM52	?	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	Y	R
MB: [LL28]	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
swON: [99EM51]	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
SK: [99em689]	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
MB: [95em413]	.	.	.	.	.	T	.	.	.	TA	.	.	.	.	.	.	.	.	.	.	A
MB: [95em403]	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
MB [95EM417]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	A
2003EM48,2003EM56	?	?	?	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
BC: [LL22]. AB: [LL36, 95EM249, 95EM254, 95EM275, 95EM298, 95EM335]. MB [95EM406, 95EM409, 95EM418]. swON: [99EM40, 99EM54, 99EM55, 99EM62, 99EM63, 99EM64, 99EM68, 99EM78, 99EM81, 99EM72, 99EM73]. nON [99EM723], NB: [LL56]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
eON: [99em744]	.	C	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	A
swON: [99EM65]	.	C	.	.	.	.	.	.	.	.	.	K	.	.	.	.	.	.	.	.	A
swON: [99EM57]	.	C	.	.	.	.	.	.	.	.	R	K	.	.	.	.	.	.	.	.	A
swON: [99EM52, 99EM53, 99EM61, 99EM66, 99EM67]. eON: [99em743]. IL : [99em141]	.	C	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A
eON: [99EM80]	.	Y	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	A

Table 17. *Lygus lineolaris*. Variation in LL56F. * = ATACC, ** = GCCAAGATAGGT. Dot indicates identity with first sample. Geographic regions: BC = British Columbia, AB = Alberta, SK = Saskatchewan, MB = Manitoba, nON = Ontario north of Algonquin Park, eON = Ontario east of Lake Simcoe, swON = Ontario west of Lake Simcoe, wQC = west Quebec, NB = New Brunswick, OH = Ohio, IA = Iowa, NE = Nebraska

Appendix A. Library and primer development methods.

A.1 DNA Isolation:

Total nucleic acids sufficiently pure for library construction was obtained using a CTAB (hexadecyltrimethyl ammonium bromide) extraction procedure. Two to four hundred mg of insect material was partitioned into six 1.5ml Eppendorph microcentrifuge tubes. Tissues were first homogenized in 100 µl of 2X CTAB extraction buffer (55 mM CTAB, 1.4 M NaCl, 50 mM Tris pH 7.5, 20 mM EDTA, 0.2% v/v beta-mercaptoethanol) using sterile disposable plastic mortars (Fisher Scientific). The final volume of each tube was adjusted to 800 µl with 2X CTAB buffer and then incubated for two hours at 65°C. Six hundred µl chloroform was added to each tissue lysate, tubes were gently pulse vortexed for one to two seconds, then centrifuged at 14000 RPM at 4°C for 10 minutes using an Eppendorph microcentrifuge. The upper aqueous phase was transferred to a new 1.5 ml tube containing 600 µl phenol/chloroform 1:1, tubes were briefly mixed by hand then incubated for 20 minutes on a rocker platform. Samples were centrifuged for 10 minutes at 14000 RPM at 4°C, the upper aqueous phase decanted and re-extracted with 600 µl chloroform. Eight hundred µl isopropanol was added to the final aqueous phase, tubes were mixed again by gentle hand agitation, then incubated at -20°C for one hour. Total nucleic acids were pelleted by centrifugation at 9000RPM at 4°C for 15 minutes, then washed three times in 1 ml 70% ethanol, centrifuging at 9000RPM at 4°C for 5 minutes for each wash. Nucleic acids were dried under vacuum, then dissolved overnight at 4°C in 50 µl low TE (1 mM Tris pH 7.5, 0.1 mM EDTA) containing one µl 1mg/ml RNase (Boehringer).

A.2 Library Construction:

Twenty mg (10 µl of extraction) total DNA was digested for 3 h at the recommended temperature in a 40 µl final reaction volume containing 15 units Tsp509 I or EcoRI restriction enzyme and 1X digest buffer 1 (New England Biolabs). The restriction digest was then extracted by the addition of 200 µl water and 200 µl phenol/chloroform 1:1, incubated on a rocker platform for 10 minutes and centrifuged for 10 minutes at 14000 RPM at 4°C. The upper aqueous phase was decanted and extracted twice with 300 µl chloroform, centrifuging for 5 minutes at 14000 RPM at 4°C for each chloroform extraction. DNA digest products were precipitated from the final aqueous phase by the addition of 20 µl 4 M sodium chloride and 600 µl absolute ethanol and incubating at -20°C for 4 hours. Precipitated DNA was pelleted by a 10 minute centrifugation at 14000 RPM at 4°C and washed three times with 1ml 70% ethanol, centrifuging for 5 minutes at 14000RPM at 4°C for each wash. DNA was dried under vacuum, then dissolved overnight at 4°C in 20 µl low TE. The concentration of DNA was quantified on a spectrophotometer and diluted accordingly in low TE to a final working concentration of 100 ng/µl.

Microsatellite libraries were generated using the Lambda ZAP II / EcoRI / CIAP-Treated Vector cloning system (Stratagene). One hundred ng of restriction-digested target DNA was ligated to 1 µg ZAP-II phage arms in a 10 µl final reaction containing 400 units T4 DNA ligase (NEB) and 1X final concentration ligation buffer. Ligation reactions were incubated at 12°C for 12 to 16 hours then stored on ice until packaging. Four µl of ligation reaction was added to one tube Gigapack III Gold packaging extract, tube contents mixed by very gentle pipetting, then incubated for 2 hours at 22°C. The packaging reaction was terminated by the addition of 500 µl SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris pH 7.5, 0.01% w/v gelatin) and 25 µl chloroform followed by a mild pulse vortexing. XL1-Blue MRF' bacterial cell line (Stratagene)

was used as host for phage growth. Host cells were prepared for phage infection in accordance to manufacturer's recommendations. A single colony of XL1-Blue MRF'cells was innoculated into 12 µl LB (Appendix 1) containing 10mM magnesium chloride and 0.2% maltose and grown for 14 hours at 37°C in a Brunswick shaker incubator at 250 RPM. Cells were then pelleted by a 10 minute centrifugation at 6000 G and resuspended in 12 ml TM (50 mM Tris pH 7.5, 10 mM MgSO4) buffer, re-incubated at 37°C for an additional 30 minutes then stored at 4°C until use. Immediately prior to phage infection, XL1-Blue MRF' bacterial host cells were diluted to an OD600nm of 0.5 to 0.6 with TM buffer. The library titre was quantified by infecting 300 µl diluted XL1-Blue MRF' cells with 100 µl of a 10-fold to 1000-fold dilution series of the SM packaging reaction in 15 mm X 100 mm glass culture tubes. After a 15 minute incubation at 37°C, 3 µl top NZY agar (Appendix 1) prewarmed to 50□C was added to the infection culture tube, the contents gently mixed, then poured onto 100 X 15 mm NZY plates (Appendix B). Plates were incubated for 14 hours at 37°C.

The titre of the library was determined and the necessary volume to plate 50,000 plaque forming units (pfu) calculated and used to infect six culture tubes containing 600 µl freshly diluted (OD600nm 0.5 to 0.6) XL1-Blue-MRF' cells. Following infection for 15 minutes at 37°C, 8 ml NZY top agarose (Appendix B) was added to the tube, the contents gently mixed, then poured onto 150 X 15 mm NZY plates. Plates were incubated at 37□C for 14 hours, then cooled to 4°C for one hour. Plaque lifts were performed in duplicate using Hybond N+ (Amersham/Pharmacia) nylon membrane, the first lift transferred for 1 minute and the replicate membrane transferred for three minutes. Membranes denatured for 5 minutes on Whatman 3MM filter paper pre-soaked in denaturing solution (0.5 N NaOH, 1.5 M NaCl), neutralized twice for 3 minutes Whatman 3MM paper pre-soaked in neutralizing solution (1.0M Tris pH 8.0, 1.5 M NaCl) and finally washed for 2 minutes on Whatman 3MM paper pre-soaked in 2X SSC solution (300 mM NaCl, 30 mM sodium citrate pH 7.0). Membranes were dried at 70°C for 30 minutes then U.V. cross-linked in a UVP CL-1000 Ultraviolet Crosslinker preset to 120 000 microjoules/cm2. Membranes were prehybridized at 30°C for three hours in prehybridization buffer (5X SSPE, 5X Denhardt's, 0.1% SDS), then hybridized overnight at 30°C following the addition of labelled oligonucleotide probes (probe preparation described below). Membranes were washed twice at room temperature for 15 minutes in 2X SSC/ 0.1% SDS, briefly blot dried with Whatman filter paper, wrapped in Saran wrap and autoradiographs generated by exposing Fugi-NIF Rx X-ray film for 48 hours at -80°C using Dupont Cronex Hi-Plus intensifying screens. Plaques containing microsatellites target inserts were pulled (4mm diameter plugs) from the plates and transferred to 1.5 ml Eppendorph tubes containing 500 µl SM buffer and 25 µl chloroform.

Cloned target inserts were excised from the purified phage in pBluescript SK- phagemid using the SOLR excision plating strain as described by Stratagene. Fifty µl secondary purified stock plug elution was added to 200 µl freshly grown and diluted XL1-Blue MRF' cells (OD600nm of 1.0) containing 1 µl ExAssist helper phage and infected for 15 minutes at 37°C. Three µl LB was added and tubes were incubated for 3 hours at 37C in a New Brunswick shaker-incubator at 250 RPM, then heated to 68°C for 20 minutes in a water bath and placed on ice until centrifugation. Tubes were centrifuged at 6000 G for 10 minutes and the supernatant decanted into Falcon 2059 centrifuge tubes and stored at 4°C. Five µl phage lysate was used to infect 200 µl SOLR cells (OD600nm 1.0) for 15 minutes at 37□C, 300 µl NZY media (Appendix 1) was added followed by 45 minutes of growth at 37°C in a shaker-incubator. Five µl of each phagemid excision was plated onto LB ampicillin plates (Appendix 1) and grown at 37°C for 14 hours. Plates were stored at 4°C until liquid culturing of bacterial colonies. One colony from each

excision was used to innoculate 3 ml LB-ampicillin (100 μg per ml) media and grown for 14 hours at 37°C in a shaker incubator at 250 RPM. Glycerol stocks for long term -80°C storage of phagemid material were prepared by aliquoting one ml liquid culture into 2ml screw cap cryovials containing 500 μl sterile glycerol. The balance of the liquid culture was poured into 1.5 ml Eppendorph tubes, cells pelleted at 14000 RPM for 1 minute, supernatant decanted and the cells suspended in 200 μl buffer (50 mM glucose, 25 mM Tris pH 7.5, 10 mM EDTA). Four hundred μl cracking solution (0.2N NaOH, 1.0% SDS) was added, tubes incubated on ice for 20 minutes, 300 μl 7.5 M ammonium acetate added, mixed and again incubated on ice for 20 minutes. Samples were then centrifuged at 14000 RPM for 10 minutes, the supernatant transferred and extracted once with 300 μl phenol chloroform 1:1 and once with 500 μl chloroform. Phagemid DNA was precipitated from the upper aqueous phase by the addition of 700 μl isopropanol and incubation at -20°C for 2 hours. DNA was pelleted at 14000RPM for 10 minutes at 4°C and washed three times in 1ml 70% ethanol. Pellets were dried in a vacuum desiccator then dissolved overnight at 4°C in 50 μl low TE and 1 μl 1mg/ml RNase.

Cloned inserts were released from the pBluescript phagmid vector by EcoRI restriction digest of 2 μl plasmid DNA and fragments separated on 1% agarose gels. Microsatellite inserts were confirmed by Unblot hybridizations as follows: gels were denatured for 30 minutes in denaturing solution (0.5 N NaOH, 1.5 M NaCl) and neutralized for 30 minutes in neutralizing solution (1.0M Tris pH 8.0, 1.5 M NaCl), rinsed in water and dried onto Whatman 3MM filter paper for 2 hours at 62°C using a model 583 gel drying unit (Bio-Rad). The dried gels were removed from the filter paper by soaking the gel/filter paper in water for 2 minutes and pealing away the filter paper. The gel was washed thoroughly to remove residual filter paper and placed immediately in 25 μl hybridization solution containing 5X SSPE/ 0.1% SDS and 32P end-labelled oligonucleotide probe. Gels were hybridized for 14 hours at 30°C, washed twice for 5 minutes in 3X SSC/ 0.1% SDS, blot dried on Whatman filter paper, wrapped in saran wrap and exposed to Fuji NIF-Rx film for 3 hours at room temperature. Clones containing microsatellite inserts ranging in size from 0.7 kilobase pairs (Kb) to 2Kb were set aside for sequencing. Microsatellite-containing inserts greater then 2Kb were digested with a variety of restriction enzymes and subsequently unblotted to confirm the release of the microsatellite loci into fragments within the 0.7Kb to 2Kb size range. Two μg of the phagemid preparation was digested with 20 units of the appropriate restriction enzyme or combination of restriction enzymes for 3 hours at the temperature and buffer conditions recommended by the manufacturer. Restriction enzyme(s) were heat inactivated at 80°C for 10 minutes and cooled on ice for 2 minutes. Restriction products were blunt ended by addition of 1 μl 2mM dNTP's (Boehringer), 0.5 μl 10X buffer 1 (NEB), 3 μl distilled water, 0.5 μl (2.5 units) DNA polymerase I, Large (Klenow) fragment (NEB) and incubating at 25°C for 15 minutes. Samples were immediately loaded onto 1% agarose gels and electrophoresed at 80 volts for 2.5 hours in 1X TBE buffer. Target bands were excised from the gel and electroeluted from the agarose gel matrix as follows: gel slice was inserted into a 3 cm strip of Spectra/Por dialysis membrane (MWCO 12-14,000: 6.4 mm diameter: Spectrum Scientific), 400 μl 0.5X TBE added followed by electrophoresis for 2 hours at 170 volts in 0.5X TBE. The fluid from the dialysis tubing was transferred to a 1.5ml eppendorph tube containing 30 μl 4M NaCl, 0.5 μl 20mg/ml glycogen (Boehringer) and 1ml absolute ethanol, mixed and incubated overnight at -20□C. DNA was pelleted by a 10 minute centrifugation at 14,000 RPM, then washed three times in 70% ethanol. Pellets were dried under vacuum to near dryness and dissolved overnight in 20 μl low TE. The concentration of DNA was determined by electrophoresing 4 μl of this purified target DNA against a low molecular weight mass ladder (Gibco-BRL). Three to 7 μl electroeluted DNA was added to 100 ng EcoRV LITMUS-28 (NEB) plasmid vector treated with calf intestinal phosphatase, 1 μl 10X NEB ligase

buffer, 0.8 µl T4 ligase enzyme (NEB) and water added to yield a final volume of 10 µl. Ligation reactions were incubated at 16°C for 14 hours. Three µl of the ligation reaction was added to 50 µl TOP-10 electrocompetent cells (Invitrogen), incubated on ice for 5 minutes then electroporated in 1 mm gap electroporation cuvettes (Molecular BioProducts) at 1.4kv and 129 ohms resistance. One ml LB was added to the cuvette, the contents transferred to a 16 X 100 mm sterile culture tube and incubated at 37°C in a shaker-incubator for 45 minutes. Fifty µl, 250 µl and 600 µl LB transformation was spread onto LB + ampicillin + IPTG + x-gal (Appendix I) plates, and grown for 14 hours at 37°C. Six white colonies from each transformation experiment were grown up in 3ml LB + ampicillin cultures, glycerol stocks made and plasmid DNA isolated as previously described. Cloned inserts were excised from the Litmus 28 vector by Xho-I / Xba-I restriction digest and confirmed to be microsatellite containing sequences by Unblot hybridization.

A.3 Oligonucleotide probe preparation:

Two oligonucleotide sequences (5' CACACACACACACAC 3' and 5' GAGAGAGAGAGAGAG 3') synthesized by Gibco BRL were diluted in HPLC grade water (Caledon) to a working concentration of 10 pmoles/µl. Oligonucleotides were end-labelled with 32P using a polynucleotide kinase (PNK) reaction. Ten pmoles of each oligonucleotide were aliquoted into a 1.5 ml eppendorph tube containing 1.5 µl 10X PNK buffer (NEB), 10 units T4 polynucleotide kinase (NEB), 5 µl gamma P32 ATP at 6000 µCi/mmol (Dupont) and 5.5 µl water. Reactions were incubated for 3 hours at 37°C then heated to 100°C for 5 minutes, cooled on ice for 2 minutes and added directly to the hybridization mix without purification of unincorporated label.

A.4 Primer Design and Testing:

Plasmid preparations containing microsatellite tracks were sequenced by a commercial service (Canadian Molecular Research Services, Ottawa) and the sequence information used to design primers using Lasergene PrimerSelect software (DNAStar Inc.). Initial primer sets were designed to cross the microsatellite track. Subsequently, additional PCR primers were designed to specifically examine sequence variability in the regions flanking the microsatellite. Sequencing reactions generally proceed more efficiently if the sequencing primers are situated away from the ends of the PCR product. Hence, additional nested primers were developed to be used in sequencing reaction. (See Figure 1 for sample primer set map and description of primer naming conventions.)

Primer sets were then tested for specificity and product yield. Approximately 100 ng of template DNA was added to 25 µl reaction mix (2.5 µl 10x reaction buffer [Boehringer Mannheim], 1.5 to 2 units Taq polymerase [Boehringer Mannheim], 2.5 mM $MgCl_2$, 40 µM each primer, 400 µM each dNTP), and amplified in a Thermolyne Amplitron II thermocycler under the following conditions: 27 cycles of denature at 94°C for 1 minute, primer annealing at 56□C for 1.5 minutes, product extension at 72°C for 1.5 minutes, followed by 8 cycles of the same conditions but with an extension time of 10 minutes. Ten µl of PCR product were loaded on a 1% agarose gel, and following electrophoresis the gel dried and probed for the presence of microsatellite-containing PCR products of the expected size. Products which hybridized were rerun on an agarose gel, the desired band excised from the gel and DNA extracted by 'freeze-squeeze' elution (gel slice frozen, then thawed and the liquid physically expressed by squeezing between sheets of paraffin film). The eluted DNA was sequenced manually using 33P-labelled

terminator chemistry (Amersham) to verify that the expected product had been obtained. PCR conditions were then tuned to maximize specificity and yield.

Assays of suitability of PCR primers for *Lygus* were performed using CTAB extractions of DNA from 6 pooled individuals of *L. lineolaris* from each of 5 sites (London ON, Durham Co. ON, Waco NE, Portage-la-Prairie MB, and Echo Bay ON) and a pooled sample of 6 individuals of *L. elisus* from Twin Falls ID.

Appendix B: Media Recipes

B.1. LB media:

10g Bactotryptone (trypticase peptone)
5g Bactoyeast extract
10g NaCl

Dissolve contents in 1L nanopure water, autoclave

- for LB-plates; add 15g agar
- for LB-ampicillin plates; following sterilization in autoclave, cool media to 50°C in water bath, add 100ug ampicillin per ml media, mix, pour plates immediately
- for LB-amp/ IPTG/ x-gal plates; take LB-amp plates, spread 40µl 100mM (20mg/ml) Isopropyl-D-thiogalactopyranoside and 60µl 20mg/ml 5-bromo-4-chloro-3-indolyl-B-D-galactoside in dimethylformamide, let plate dry for approximately 30 minutes prior to plating transformants
- for LB-top agar/agarose; add 7g agar or agarose

B.2. NZY media:

10g NZ amine (casein enzymatic hydrolysate)
5g Bactoyeast extract
5g NaCl
2g $MgSO_4$-$7H_2O$

Dissolve contents in 1L nanopure water, autoclave

- for NZY-plates; add 20g agar
- for NZY-top agar/agarose; add 7g agar or agarose

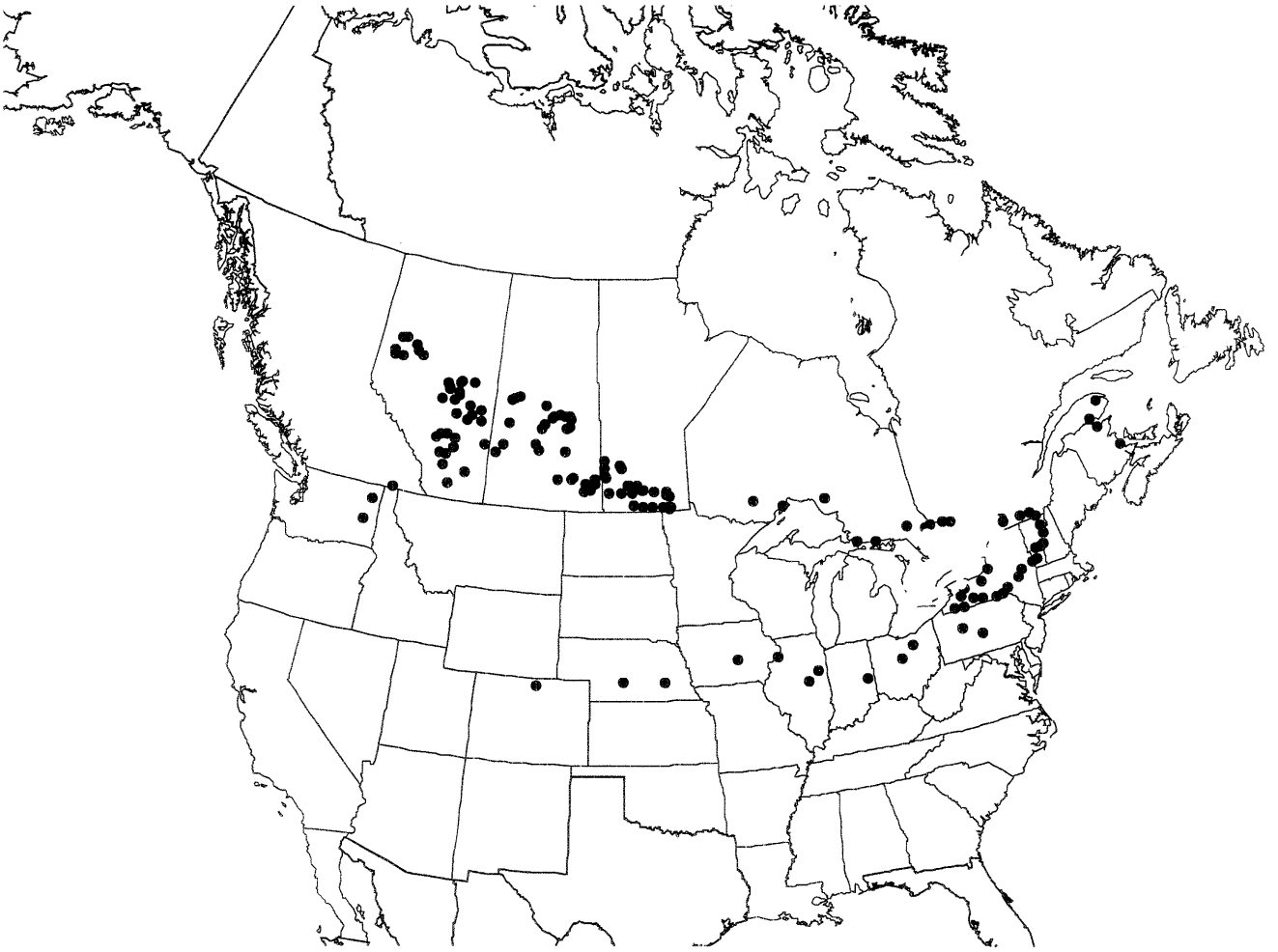

Figure 1. Collection localities, except southern Ontario and west Quebec (see Fig. 2).

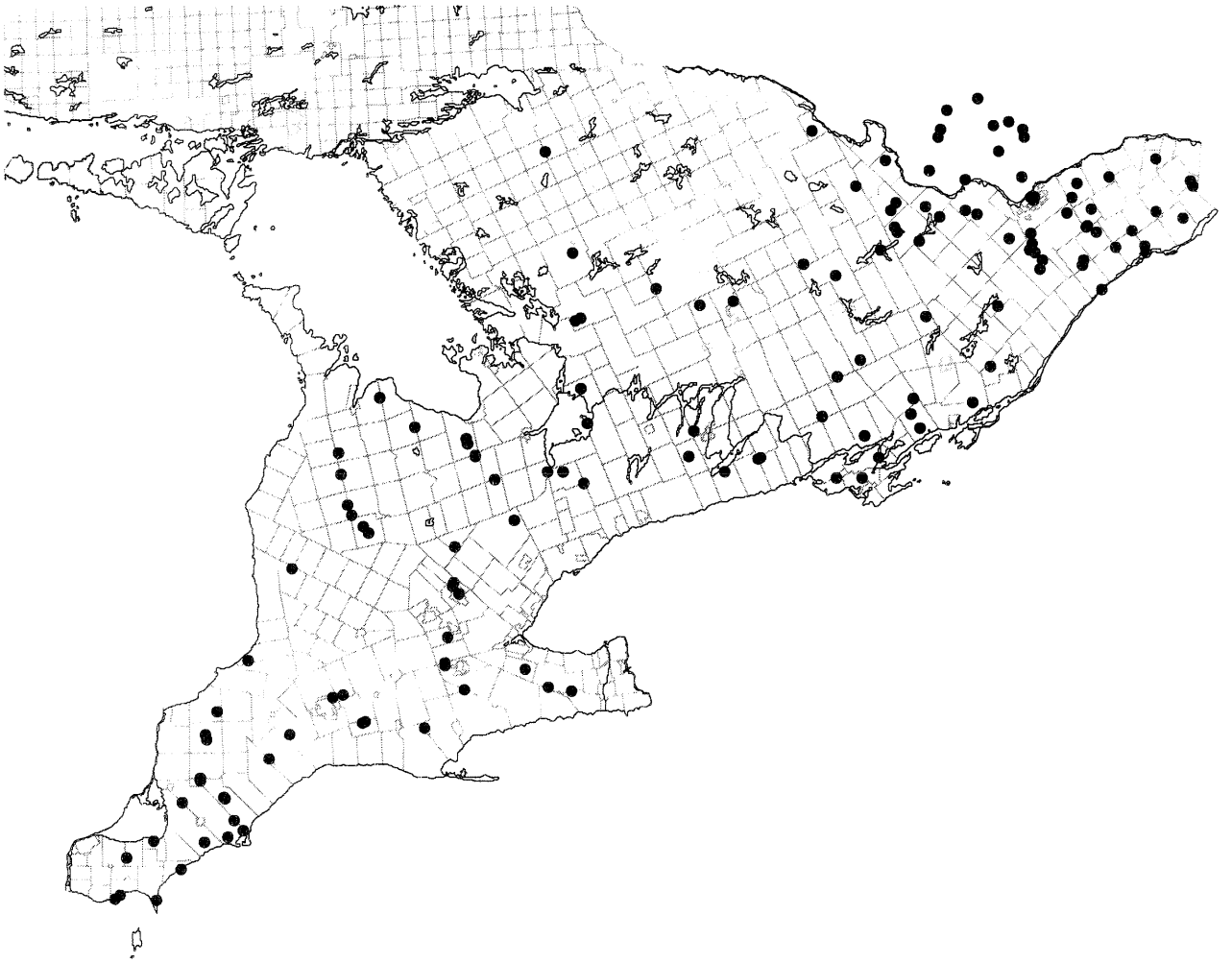

Figure 2. Collection localities in southern Ontario and west Quebec.

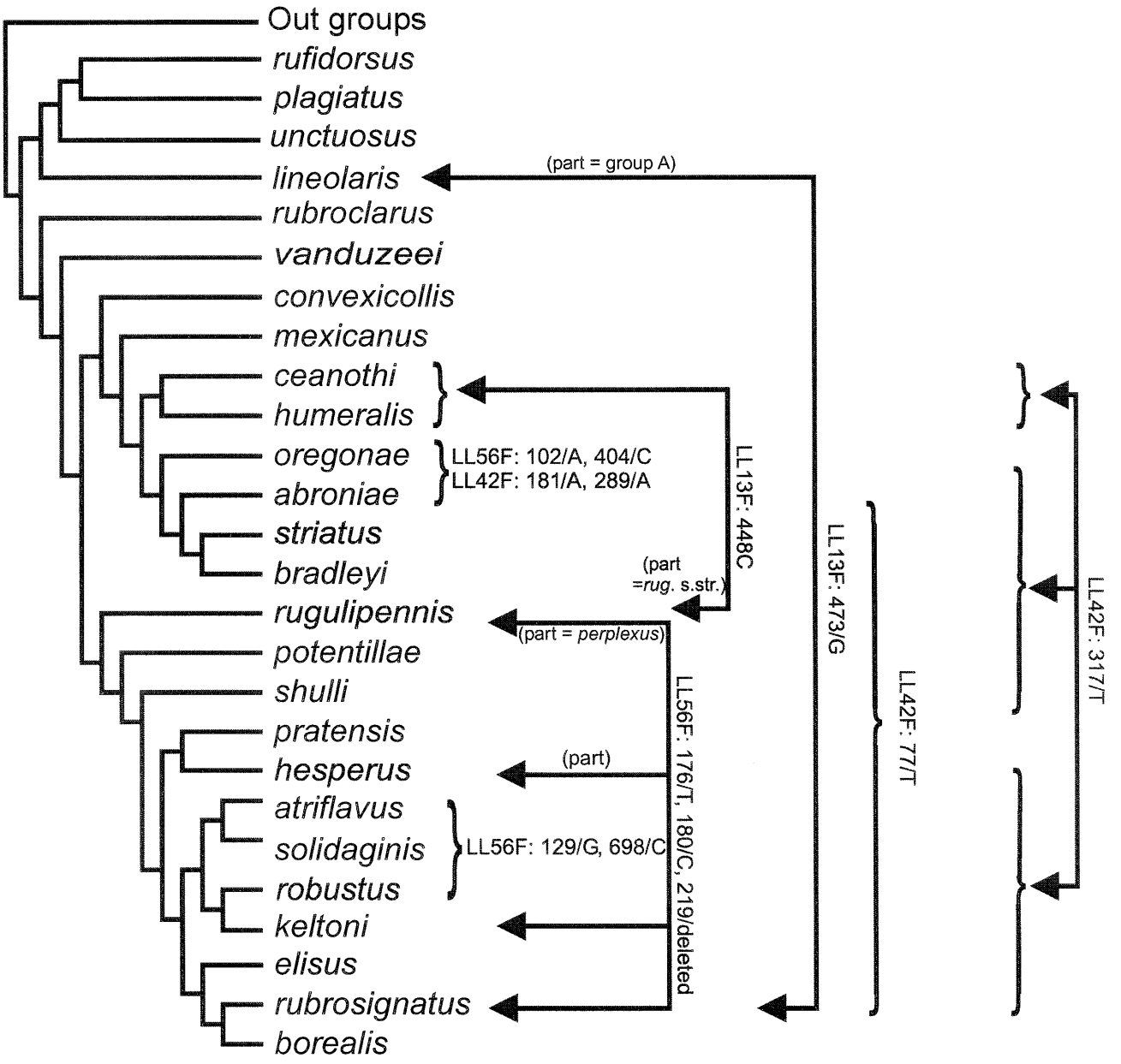

Figure 3. Derived state of informative characters (shared changes) for all loci mapped onto a tree based on morphological data (adapted from Schwartz and Foottit, 1998). Character states are indicated as "Locus name: position/derived state".