ORIGINAL PAPER

Population-based resequencing analysis of improved wheat germplasm at wheat leaf rust resistance locus *Lr21*

Yong-Bi Fu · Gregory W. Peterson · Brent D. McCallum · Li Huang

Received: 28 September 2009 / Accepted: 22 February 2010 / Published online: 10 March 2010 © Springer-Verlag 2010

Abstract Little is known about the genetic impacts of modern plant breeding on specific breeding target loci. Resequencing cloned genes can identify all mutations in single genes for population-based analyses of genetic changes in improved gene pools. Ninety-five wheat cultivars released in Canada from 1845 to 2004 were sequenced at the wheat leaf rust resistance locus *Lr21*. Characterization of the DNA fragment of length 4,071 bp, covering the $Lr21$ gene from -92 to $+4,261$, revealed 13 SNPs, four indels, 10 haplotypes, and 4 major haplotype groups. A new SCAR marker was developed to identify the resistant haplotype and haplotype groups. Non-synonymous polymorphic sites and haplotype numbers were increased over the 100 years of wheat breeding. Nucleotide diversity of the wheat cultivars was gradually reduced from 1845 to 1993 and increased after the release of the first *Lr21* wheat cultivar AC Cora in 1994. Positive selection measured with

Communicated by M. Sorrells.

Electronic supplementary material The online version of this article (doi[:10.1007/s00122-010-1308-7](http://dx.doi.org/10.1007/s00122-010-1308-7)) contains supplementary material, which is available to authorized users.

Y.-B. Fu $(\boxtimes) \cdot G$. W. Peterson

Plant Gene Resources of Canada,

Saskatoon Research Centre, Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N 0X2, Canada e-mail: yong-bi.fu@agr.gc.ca

B. D. McCallum

Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Road, Winnipeg, MB R3T 2M9, Canada

L. Huang

Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, MT 59717-3150, USA Tajima's *D* was observed in the cultivars released before 1935. At least two recombination events were inferred in those cultivars released before 1993. Linkage disequilibrium at the locus was decreased over time. These findings demonstrate not only the effectiveness of the wheat breeding in the improvement of leaf rust resistance, but also are useful to understand the genetic influences of a long-term artificial selection on individual loci.

Introduction

Recent years have seen an increasing interest in molecular assessments of genetic diversity changes in existing gene pools of major agricultural crops (e.g., Donini et al. [2000](#page-9-0); Fu et al. [2003;](#page-9-1) Hyten et al. [2006;](#page-10-0) Malysheva-Otta et al. [2007;](#page-10-1) Fu and Somers [2009](#page-9-2)). Such assessments are not only useful to understand the effectiveness of plant breeding in the modification of plant genomes, but also help to address a long-standing concern that modern plant breeding reduces crop genetic diversity, which may have consequences for the vulnerability of crops to changes in pests, diseases, climate and agricultural practices (National Research Council [1972](#page-10-2); Tripp [1996](#page-10-3); Borlaug [2007](#page-9-3)). In general, the genome-wide reduction of crop genetic diversity accompanying genetic improvement over time is minor, but allelic reduction at individual chromosomal segments is substantial (Fu [2006\)](#page-9-4). However, little attention has been paid to molecular assessments of the genetic impacts of modern plant breeding on specific breeding target loci.

Resequencing cloned genes can identify all mutations in a gene that are present within the sampled individuals. With an increasing number of cloned genes available, this approach should present a new opportunity to characterize

breeding target genes and assess genetic changes at these loci under selective breeding through population-based analyses of nucleotide sequence diversity. Resequencing of candidate genes within diverse populations has shown the potential in understanding complex human traits (Topol and Frazer [2007](#page-10-4)). Resequencing-based studies have been used, for example, to assess vernalization response and pathogen resistance in *Arabidopsis* (Aranzana et al. [2005](#page-9-5)), flowering time adaptation in barley (Jones et al. [2008](#page-10-5)), and genetic diversity in wild avocado (Chen et al. [2008\)](#page-9-6). However, the potential of this approach has not been explored in improved crop gene pools established from long-term breeding efforts.

Leaf rust, caused by the fungal pathogen *Puccinia triticina* Eriks. (Kolmer [1996\)](#page-10-6), is one of the most important diseases of wheat (*Triticum aestivum* L.) worldwide, causing premature defoliation that results in as much as a 40% yield loss (McIntosh et al. [1995](#page-10-7)). The long-term strategy for the control of this disease is the utilization of genetic resistance by gene pyramiding. To date 67 wheat leaf rust resistance genes have been reported (e.g., see Marais et al. [2008\)](#page-10-8) and many of them have been tagged by molecular and cytological markers (e.g., Huang and Gill [2001](#page-9-7)) to facilitate leaf rust resistance breeding. Four of these genes have been cloned: *Lr1* (Cloutier et al. [2007](#page-9-8)), *Lr10* (Feuillet et al. [2003\)](#page-9-9), *Lr21* (Huang et al. [2003](#page-10-9)) and *Lr34* (Krattinger et al. [2009\)](#page-10-10). Both *Lr1* and *Lr10* are ineffective against some virulence phenotypes of *P. triticina* but *Lr21* and *Lr34* remain effective against all isolates in North America. *Lr21* was first transferred from *Aegilops tauschii* (Coss.) accession TA1599 to the stem rust resistant wheat cultivar Thatcher in 1970s (Rowland and Kerber [1974;](#page-10-11) McIntosh et al. [1995\)](#page-10-7) and has been available for breeding since the 1970s. However, the first developed cultivar to carry fully functional *Lr21* gene was AC Cora, which was released in 1994 (McCallum et al. [2007](#page-10-12)). *Lr21* is a gene of length 4,318 bp at a simple (single-copy) locus on the chromosome 1DS, encodes a nucleotide-binding site-leucine-rich repeats (NBS-LRR) protein of 1,080 amino acids, and confers resistance to all known *P. triticina* races (Huang et al. [2003](#page-10-9), [2009](#page-10-13)). Thus, this cloned gene provides a unique re-sequencing opportunity to assess the genetic changes in improved Canadian wheat gene pool.

Canadian wheat breeding began in 1886, has so far released hundreds of cultivars, and has generated a significant impact on Canadian agriculture (DePauw et al. [1995](#page-9-10)). Breeding targets have changed from adaptation and quality before 1940, resistance to biotic and abiotic stresses such as rust from 1940 to 1990, to end-use quality such as increased grain protein after 1990 (Neatby [1942](#page-10-14); Morrison [1960](#page-10-15); McCallum and DePauw [2008](#page-10-16)). The accompanying breeding methods range from introduction, mass selection, hybridization, backcrossing, to marker-assisted selection (Slinkard and Knott [1995;](#page-10-17) McCallum and DePauw [2008](#page-10-16)). These practices have generated considerable impacts on the genome of improved wheat cultivars (Fu et al. [2005](#page-9-11), [2006](#page-9-12)). A genome-wide reduction of genetic diversity was observed in 75 Canadian hard red spring (HRS) wheat cultivars released from 1845 to 2004 (Fu and Somers [2009](#page-9-2)). However, little is known about the genetic impacts on specific loci of wheat breeding targets.

Historically, the Canadian breeding efforts for rust resistance were made largely toward stem rust, which is caused by *Puccinia graminis* Pers.: Pers. f. sp. *tritici* Eriks. & E. Henn. (McCallum and DePauw [2008\)](#page-10-16). Thatcher was the first significant stem rust resistant cultivar, and was grown extensively from 1939 until the early 1970s. However, it was very susceptible to leaf rust. Over time, improved resistance to both stem and leaf rust was achieved with the release of cultivars with additional genes for resistance, primarily *Sr2*, *Sr6*, *Sr7a*, *Sr9b*, *Lr13*, *Lr14a*, *Lr16*, *Lr21*, *Lr22a* and *Lr34*. Overall, genetic resistance has adequately controlled stem rust but leaf rust continues to cause significant losses, as changes in the *P. triticina* population gradually reduced the effectiveness of resistance genes such as *Lr13*, *Lr14a* and *Lr16*. While some resistance genes incorporated since the 1970s such as *Lr21* and *Lr34* remain largely effective, research efforts have been directed in recent years to combine additional resistance genes such as *Lr18*, *Lr35*, *Lr46* and *Lr52* into new breeding lines (McCallum et al. [2007\)](#page-10-12).

The objectives of this study were to (1) characterize the nucleotide sequence variation at the wheat leaf rust resistance locus *Lr21* in 95 wheat cultivars released in Canada from 1845 to 2004, (2) develop a SCAR marker for identification of *Lr21* haplotypes, and (3) assess the genetic changes at the locus over various breeding periods.

Materials and methods

Plant materials

Ninety-five wheat cultivars (Table [1\)](#page-2-0) were selected from the wheat germplasm collection maintained at Plant Gene Resources of Canada, Saskatoon Research Centre. The selected cultivars consisted of three major types of wheat cultivars [75 HRS, 9 soft white spring (SWS), 11 winter] and are representative of the long-term Canadian wheat breeding efforts. More HRS wheat cultivars were selected as the HRS wheat breeding reflected the major effort of Canadian wheat breeding over the past century. Additional descriptions of these cultivars are given in supplemental Table S1; they include the origin of release, seedbank and GenBank accession labels.

Table 1 List of 95 wheat cultivars with the release year, wheat type, inferred haplotype, field resistance rating, and haplotype grouping

DNA extraction

Five seeds of each cultivar were randomly selected from the PGRC wheat collection and grown in the greenhouse at the Saskatoon Research Centre. Young leaves were individually collected from five 5-day-old seedlings of each cultivar, freeze-dried (in a Labconco Freeze Dry System for 3–5 days), and stored at -80° C. Dry leaves from one individual sample of each cultivar were finely chopped and ground to a fine powder in a 2-ml microcentrifuge tube with two 3-mm glass beads on a horizontal shaker. Genomic DNA was extracted using a DNeasy Plant Mini Kit (Qiagen Inc., Mississauga, ON, Canada) according to the manufacturer's directions. Extracted DNA was quantified by fluorimetry using Hoechst 33258 stain (Sigma Chemical Co., St. Louis, MO, USA), followed by dilution to 25 ng μ l⁻¹.

Lr21 sequencing

A 4 kb region of $Lr21$ was amplified using primers $3F (5' - T)$ GC TTC CGT GTT TGA TCC TGC CAC TT-3') and 3R (5'-AAC CAG CCT GCC ACA ACC TCT TTA CT-3') under the following conditions: 1 U Phusion polymerase (Finzyme, New England Biolabs, Pickering, ON, Canada) in $1 \times$ Phusion HF buffer containing 1.5 mM MgCl₂, 0.2 mM each dNTPs, 0.25μ M each primer, 25 ng of template genomic DNA in a final volume of 50 µl. A two-step PCR was carried out at 94°C 3 min, 30 cycles of 94°C 10 s, 72°C 3 min, followed by 72°C 10 min, and cooling to 20°C. Two microliters of the PCR product was used as a template to reamplify two fragments overlapping by 100 bp using primer pairs 3CL22F (5'-GCT TGG AAA TGC TGT CCA TGC TGA-3') and 22R (5'-CGT TTG CCA CAC AAT GCT TCC TTC-3'), and 4CL20R (5'-TGT TCG TCA TCT TCG TTC GTC GGT-3') and 20F (5'-CGG AGA TAG TCA CTC CAA TAT TTC AGA TCG-3') under the following conditions: 1 U *Taq* polymerase (New England Biolabs, Pickering, ON, Canada), $1 \times$ Thermopol Buffer with $2 \text{ mM } MgSO_4$ (New England Biolabs, Pickering, ON, Canada), 0.2 mM each dNTP, $0.25 \mu \text{M}$ each primer, in a final volume of 50 μ l. The PCR was carried out at 94°C 3 min, 30 cycles of 94°C 10 s, 59°C 20 s, 72°C 2 min 30 s followed by 72°C 5 min, and cooling to 20°C.

The PCR product of 3CL22F-22R was sequenced using primers: 3CL22F, 3seq8F (5'-ACG GAG GAG CTC TTT GTG TCG TTT-3'), and 3seq16F (5'-AGA ACT GCT AGC AAG CGG AAG GTA-3') at the DNA sequencing laboratory, Plant Biotechnology Institute, National Research Council of Canada, Saskatoon. The PCR product of 4CL20R-20F was sequenced using primers: 4CL20R, 4seq8R (5'-TCC TTA GCC TGA AGA AGG GT-3'), and 4seq16R (5--AGC TGC TGA TAG CTC CAC CAA AGA-3'). Additional re-sequencing was independently

made three times for five relevant cultivars to verify all newly detected SNPs.

SCAR marker for *Lr21* genotyping

Primers F1 (5'-ACATCCAATTTCGTCGCCCACT-3') and R1 (5--CTGCAACTTCCGCATTGCCATT) were used to amplify a 2.4 kb fragment of the *Lr21* gene covering the functional region from the second to fourth indels. The PCR cocktail consisted of $1 \times$ New England Biolabs (Pickering, ON, Canada) Thermopol Buffer containing 1.5 mM MgSO4, 1 U of *Taq* polymerase (New England Biolabs) 0.2 mM each dNTP, $0.2 \mu M$ each forward and reverse primer, and 100 ng of genomic DNA in a final volume of 25μ . The fragment was amplified under the following conditions: 94°C 3 min, 30 cycles of 94°C 10 s, 59°C 15 s, and 72°C 2 min followed by an extension step of 72°C 5 min. Upon completion of the PCR reaction 10 U of *Bcl*I restriction enzyme (New England Biolabs) was added to the PCR cocktail directly and incubated for 2 h at 50°C. The resulting digest was separated on a 1.5% agarose gel for 3 h at 100 V, stained with 0.5 mg/l ethidium bromide for 10 min and photographed. To confirm the effectiveness of the developed SCAR marker, seven representative *Lr21* sequenced cultivars and seven additional wheat cultivars released in the last decade were tested. The additional cultivars were CDC Alsask, Fieldstar, Kane, Unity, Helios, Infinity, and Somerset. Application of the SCAR marker was also made to eight barley cultivars, eight cultivated oat cultivars, and six accessions representing four wild oat species (*Avena sterilis* L., *A. fatua* L., *A. nudisativa* L., and *A. byzantina* K. Koch).

Field test of leaf rust resistance

Leaf rust resistance was assessed in inoculated and natu-rally infected field trials (Table [1\)](#page-2-0). In the inoculated trials, spreader rows of leaf rust susceptible lines were placed every alternating third row to increase the leaf rust inoculum. The inoculum used throughout the phenotyping process was an epidemic mixture of an equal proportion of the most prominent isolates collected during the annual leaf rust disease surveys in the year prior to the trial (e.g., see McCallum and Seto-Goh [2008](#page-10-18)). Spreader rows were first inoculated at the stem elongation phase. The inoculum mixture was a suspension of 1.5 g of urediospores/l of light mineral oil (Bayol Esso Canada, Toronto, ON, Canada). A hand held low volume herbicide applicator was used to apply the inoculum to the spreader rows only. Terminal severity of leaf rust on the flag leaves was determined using the modified Cobb Scale ($%$ infection of flag leaf) (Peterson et al. [1948](#page-10-19)). The cultivars were assigned resistance levels based on relative resistance and on the following scale for

inoculated nurseries $\langle 10\% = R$ or resistant, $11-30\% = MR$ or moderately resistant, $31-40\% = I$ or intermediate, $41 60\% = MS$ or moderately susceptible, $>60\% = S$ or susceptible.

Data analysis

All sequencing products were assembled with Vector NTI Suite's ContigExpress v9.0.0 (Invitrogen, Carlsbad, CA, USA) and aligned using MUSCLE v3.6 (Edgar [2004](#page-9-13)). All aligned sequences have been deposited in the GenBank database and their accession numbers are listed in supplemental Table S1. Population genetic analyses of aligned DNA sequences were performed using DnaSP program (Librado and Rozas [2009](#page-10-20)). Several measures of sequence variation were obtained, including the number of segregating sites, haplotype number, nucleotide diversity $(\pi;$ Tajima [1983](#page-10-21)), the signal of selection (i.e., deviation from neutrality, *D*; Tajima [1989\)](#page-10-22), the frequency of recombination (i.e., the minimum number of recombination events, R_m ; Hudson and Kaplan [1985\)](#page-10-23), and the extent of linkage disequilibrium (LD) (i.e., the congruence among adjacent segregating sites, Wall's *B*; Wall [1999](#page-10-24)). Plots were made of the resul- $\tan \frac{R^2}{\pi}$ for each pair of sites against the number of nucleotides that separated the pair, and a logarithmic curve was fitted into the plots in Microsoft Excel.

Haplotype analyses with and without gaps were made using the DnaSP program. The positions of SNPs and indels for each haplotype were generated. The first instance of a haplotype within the wheat gene pool was inferred with the release year of the oldest cultivars with the haplotype. Evolutionary relationships of haplotypes were inferred with a neighbor-joining method, excluding gaps, using the MEGA program (Tamura et al. [2007](#page-10-25)) and a circular tree of the wheat cultivars was generated. The haplotype groups were established, and haplotypes or haplotype groups were associated with the field resistance ratings of the tested cultivars. Frequency distributions of haplotype groups for cultivars of the four breeding periods defined below were calculated and tested for association between haplotypes and breeding periods using likelihood ratio chi-square test in Proc FREQ $(SAS Institute 2008)$ $(SAS Institute 2008)$ and for differences among breeding periods using a random permutation procedure (Fu [2010\)](#page-9-14).

Comparative population genetic analyses were also made using the DnaSP program for 12 sub-groups of wheat cultivars (all assayed cultivars, winter wheat, SWS wheat, HRS wheat, four breeding period-specific groups of HRS wheat, and four breeding period-specific groups of all assayed cultivars). Four breeding periods (before 1935, 1935–1958, 1959–1993, after 1993) were defined to reflect the four major breeding efforts in improvement of stem and leaf rust resistances in the Canadian wheat cultivars (McCallum and DePauw [2008](#page-10-16)).

Results

The resequencing effort generated a DNA fragment of length 4,071 bp, covering the $Lr21$ gene from -92 to +4,261. Alignments of the generated sequences from the *Lr21* locus revealed the same gene structure as previously reported (Huang et al. [2003](#page-10-9)), including amino-terminus, followed by an intron, NBS, LRR, and carboxy-terminus with another intron (Fig. [1](#page-4-0)). Three previously reported indels $(2 \text{ bp in } A$ -terminus, 88 or 105 bp in the first intron, and 1 bp in the NBS region) and one new indel in the LRR region was detected (Fig. [1](#page-4-0)). The new indel of 86 bp was the same as indels 12, 13, and 14 in *Ae. tauschii* (Huang et al. [2009\)](#page-10-13), but it was not reported in hexaploid wheat. The second indel displayed variation in length (either 88 or 105 bp), and the fourth indel consisted of three parts of different lengths (Fig. 2).

The characterization of this DNA fragment for 95 wheat cultivars revealed a total of 13 SNPs (Fig. [2\)](#page-5-0). These SNPs were distributed over the three gene regions, including

Fig. 1 Applied primer pairs and gene structure of wheat *Lr21* gene. The *top row* shows the SCAR marker primers designed to detect the resistant haplotype and haplotype groups, and the *second row* displays the primer pairs used to sequence the cultivars. The gene structure

identified by Huang et al. (2003) (2003) includes promoter + 5' UTR, aminoterminus with intron, non-binding sequence, leucine-rich repeat, followed by carboxy-terminus. Below the gene structure are two cultivars with different haplotypes carrying four indels

| | | --Amino terminus-- | | | | | ---NBS--- | | | -LRR· | | | | | | | | | | | | | | |
|----------------------|----------------|--------------------|----------------|-----|----------------|----------------|-----------|----|--------|-------|------|----------------|----------------|------|--------------------------|------|--------|-----------------|------|----------|----------|-------|----------|-----------|
| SNPs | | | $\overline{2}$ | | 3 | | | | 4 | 5 | 6 | $\overline{7}$ | | | | 8 | 9 | 10 ¹ | | | 11 12 13 | | | |
| Indels | | | | 1 | | $\overline{2}$ | $2*$ | 3 | | | | | $\overline{4}$ | | $4*4**$ | | | | | | | | | Haplotype |
| Length (bp) | | | | 2 | | 88 | 17 | 1 | | | | | 4 | 9 | 73 | | | | | | | | features | |
| Position | | CSS | 212 | Γ9, | ∞ 43 | \times | \times | نى | 1864 | 52 Li | 2797 | 3193 | 3195 | 3244 | 3260 | 3285 | 3691 | 3736 | 3827 | 390 ᠗ | 392 | Count | Group | Year |
| | | | | | | | | | | | | | | | | | | | | | | | | |
| Lr21-AC Cora | h1 | | | | | | | А | G | G | | А | A | G | G | G | A | G | G | | G | 3 | | G1 1994 |
| Ir21-Red Fife | h2 | G | | | | | | | * | А | * | \ast | $*$ | $*$ | * | | \ast | $*$ | * | * | | 9 | | G2 1845 |
| <i>lr21</i> -Fielder | h ₃ | * | * | А | | | | | \ast | A | * | \ast | \ast | $*$ | * | | \ast | $*$ | * | * | | 12 | | G2 1976 |
| Ir21-Regent | h4 | * | * | А | | | | | * | A | * | \ast | $*$ | $*$ | * | | \ast | $*$ | * | * | | 1 | | G2 1939 |
| Ir21-CDC Buteo | h ₅ | * | * | Α | | | | | | А | G | | | | ۰ | | G | $*$ | * | * | | 1 | | G4 2002 |
| Ir21-Ladoga | h ₆ | G | | | | | | | | | G | | | ۰ | ۰ | | G | $*$ | А | C | | 14 | | G4 1887 |
| Ir21-Stanley | h7 | G | | | | | | | | А | G | | | | ٠ | | G | C | А | | | 20 | | G4 1893 |
| Ir21-AC Radiant h8 | | G | | | * | | | | * | А | * | | | | $\overline{}$ | | G | $*$ | Α | | | 1 | | G3 2004 |
| Ir21-Thatcher | h9 | * | * | А | \ast | | | | * | А | * | | | | $\overline{}$ | | G | C | Α | | | 31 | | G3 1935 |
| Ir21-Dawson's | h10 | * | * | А | \ast | | | | \ast | A | * | | | | | | G | $*$ | Α | | | 3 | | G3 1881 |

Fig. 2 Nucleotide polymorphism of 10 haplotypes inferred from 95 wheat cultivars. Each haplotype is labeled with its oldest cultivar and its features (count, group and the earliest year in the improved gene pool) are given in the last three columns. The position of a polymorphic site starts from the transcribed site and relative positions of major gene

regions are also given above. The *X*'s represent deletions in *Lr21*. Thirteen SNPs are labeled, and four indels along with indel length (bp) are displayed. The newly detected SNPs and indel in hexaploid wheat are *highlighted*. Variations in two large indels (#2 and #4) are shown with *asterisk*(*s*)

introns, but only two of them (no. 8 and no. 10) were newly detected in hexaploid wheat. One SNP (no. 8) was found in the fourth indel. Two SNPs (no. 11 and no. 12) in the LRR regions were synonymous, while the other SNPs were nonsynonymous. Roughly, three SNPs were detected for every thousand base pairs. Haplotype analysis considering or without gaps revealed ten haplotypes (Fig. [2\)](#page-5-0). The frequencies of these ten haplotypes ranged from 1 (for h4, h5, and h8) to 31 (h9) (Fig. [2](#page-5-0)).

The neighbor-joining clustering of the 95 wheat cultivars revealed the evolutionary relationships among ten haplotypes (Fig. [3](#page-6-0)). Roughly, the clustering corresponds to four major haplotype groups. The first distinct group with three cultivars consisted of haplotype 1. The second group with 22 wheat cultivars was closely related to the first group and consisted of haplotypes 2–4. The third and fourth groups each with 35 cultivars consisted of haplotypes 8–10 and haplotypes 5–7, respectively. These four groups of haplotypes were named G1, G2, G3, and G4, respectively (Fig. [2;](#page-5-0) Table [1\)](#page-2-0), for ease of evaluating haplotype changes. Associating these haplotype groups with the field leaf rust resistance ratings of the tested cultivars (Table [1](#page-2-0)) confirmed the leaf rust resistance conferred by the haplotype 1 of the *Lr21* locus (i.e., present in AC Cora, McKenzie, and Lovitt), and revealed the leaf rust resistance conditioned by genes at other resistance loci, rather than at the *Lr21* locus. For example, the group G2 has one cultivar (i.e., Pacific), G3 has three cultivars (i.e., Pasqua, AC Minto, and Lillian), and G4 has two cultivars (i.e., Laura and Invader) with leaf rust resistance; all of these resistances were not conditioned by the haplotype 1 of the *Lr21* locus.

A SCAR marker was developed to identify four haplotype groups and tested with four known haplotype groups and cultivars with unknown haplotype groups. The patterns of restriction fragment length polymorphism (Fig. [4\)](#page-6-1) show four haplotype groups are distinguishable, particularly the G1 haplotype group, for the *Lr21* allele. Seven cultivars with known haplotype groups confirmed the expectation for the SCAR marker. Four of the seven cultivars without *Lr21* sequencing (CDC Alsask, Fieldstar, Kane and Unity) were found to be the G1 haplotype group, while Helios, Somerset and Infinity were $G2$, $G3$, and $G4$ haplotype groups, respectively (Fig. [4\)](#page-6-1).

Application of this SCAR marker to barley, oat, and wild oat species revealed the presence of G2 haplotypes in barley, G2 and G3 haplotypes in cultivated oat, G2 haplotypes in *A. fatua*, and none in the three wild oat species (results not shown). This result confirmed that the haplotype group G2 was the oldest in hexaploid wheat. Comparison of the first instances of the ten haplotypes within the improved wheat gene pool (Fig. [2](#page-5-0)) revealed that the haplotypes 1, 4, 5, and 8 were newly derived ones, while the others were the founder haplotypes in the gene pool. The haplotype 2 represented by Red Fife was the common founder haplotype in hexaploid wheat.

Assessments of ten haplotypes over the four breeding periods revealed that haplotypes 6 and 7 occurred in all the breeding periods, while the other haplotypes were present only in some breeding periods (Table [2](#page-7-0)). Haplotype 1 occurred only in the recent breeding period, while haplotype 9 was missing in the first breeding period. Likelihood ratio test demonstrates a significant association (*P* < 0.0001) between haplotypes and breeding periods. Tracking the dominant haplotypes 9 and 7, through known pedigrees, clearly revealed the genetic influences of the early dominant cultivars Thatcher and Marquis. The

Fig. 3 Neighbor-joining clustering of 95 wheat cultivars based on *Lr21* DNA sequences, illustrating the evolutionary relationships among identified haplotypes and the four major groups of the wheat cultivars. Each cultivar is labeled with a short form, followed by a *number* representing haplotype and a letter *s* for soft spring wheat and *w* for winter wheat (or no letter for hard red spring wheat)

Fig. 4 The restriction fragment length polymorphism of the *Lr21* functional region amplified by the SCAR primer pair, illustrating the resistant haplotype and haplotype groups. *Lr21* gene sequences are available for the first seven cultivars shown with their haplotype groups, but not for the last seven testing cultivars. The four testing cultivars (CDC Alsask, Fieldstar, Kane and Unity) share the same pattern of variation with *Lr21* gene sequenced cultivars AC Cora, McKensize and Lovitt, confirming the presence of the resistant haplotype in these tested cultivars

number of haplotypes increased from four in the first two breeding periods to eight in the last breeding period (Table [2\)](#page-7-0). Such a change was statistically significant $(P < 0.025)$, based on the random permutation assessment. The dominant haplotype group was changed from G3, G4, to G2 over the four breeding periods. A similar pattern of variation was observed for HRS wheats, if G1 and G2 were combined.

The resequencing also allowed for an assessment of nucleotide sequence variation at the *Lr21* locus over various breeding periods. The nucleotide diversity analyses revealed variable patterns of variation in the number of haplotypes, the number of segregating (or polymorphic) sites, and the nucleotide diversity (Table [3](#page-7-1)). More segregating sites were found in HRS wheat, particularly for the cultivars released after 1993. Obviously, non-synonymous polymorphic sites were increased over the breeding periods. More haplotypes were observed in HRS than in SWS and winter wheat. Haplotype numbers also increased over the 100 years of wheat breeding.

The assessment of the changes in nucleotide diversity $(\pi \times 10^3)$ of the 95 wheat cultivars revealed a decrease from 0.90 before 1935 to 0.72 during 1959–1993 and an increase (1.23) after the release of the first $Lr21$ cultivar AC Cora in 1994. Similar patterns of variation were observed for HRS wheat released over the four breeding periods. However, this pattern of variation is not consistent with the increase of polymorphic sites from 7 to 12 over the same time period. Positive selection was observed in HRS wheat (Tajima's $D = 2.247$), not in the other two market classes (Table [3\)](#page-7-1). This positive selection occurred mainly in the highlighted with bold

| Table 2 Percentages of wheat cultivars in each breeding period | | Breeding period/sample size ^a | | | | | | | | | | |
|--|----------------|--|--------------|--------------|---------------|--|--|--|--|--|--|--|
| for $Lr21$ haplotypes or haplotype groups | | Before 1935/23 | 1935-1958/15 | 1959-1993/34 | After 1993/23 | | | | | | | |
| | Haplotype | | | | | | | | | | | |
| | h1 | | | | 13.0 | | | | | | | |
| | h2 | 34.8 | | 2.9 | | | | | | | | |
| | h3 | | | 8.8 | 39.1 | | | | | | | |
| | h4 | | 6.7 | | | | | | | | | |
| | h ₅ | | | | 4.3 | | | | | | | |
| | h ₆ | 34.8 | 6.7 | 8.8 | 8.7 | | | | | | | |
| | h7 | 26.1 | 60.0 | 8.8 | 8.7 | | | | | | | |
| | h ₈ | | | | 4.3 | | | | | | | |
| | h ₉ | | 26.7 | 64.7 | 21.7 | | | | | | | |
| | h10 | 4.3 | | 5.9 | | | | | | | | |
| | Group | | | | | | | | | | | |
| | G1 | | | | 13.0 | | | | | | | |
| | G ₂ | 34.8 | 6.7 | 11.8 | 39.1 | | | | | | | |
| ^a The dominant haplotype or | G ₃ | 60.9 | 66.7 | 17.6 | 21.7 | | | | | | | |
| group for each breeding period is highlighted with hold | G4 | 4.3 | 26.7 | 70.6 | 26.1 | | | | | | | |

Table 3 Summary of measures of nucleotide sequence variation at the wheat *Lr21* locus in 95 wheat cultivars

^a Synonymous and non-synonymous polymorphic sites are given in parentheses, respectively

^b Haplotype number inferred without gaps

^c Significance of test: $^{#}P \approx 0.05;$ $^{*}P < 0.05;$ $^{**}P < 0.01$

HRS wheat before 1935 and after 1993 (Tajima's $D = 2.501$ and 1.758, respectively). However, negative Tajima's *Ds*, although not statistically significant, were observed for the cultivars released from 1935 to 1993.

The recombination assessment revealed at least two recombination events at the *Lr21* locus and a trend of increase in genetic exchange for the cultivars released after

1959 (Table [3\)](#page-7-1). One recombination event occurred in the winter wheat cultivars released from 1935 to 1958 and another recombination happened in the HRS wheat cultivars released from 1959 to 1993. LD was lower in HRS wheat (Wall's $B = 0.273$) than in winter and SWS wheat (Wall's $B = 0.75$ and 0.667, respectively). Considering the HRS wheat alone, LD was decreased over the four breeding

periods; a similar pattern was observed for all the 95 wheat cultivars (Table [3](#page-7-1)). The decay of LD was more severe for the wheat cultivars after the introduction of the *Lr21* gene into the gene pool, with a sharp drop of base pair distance in the range of 200–400 bp (Fig. [5](#page-8-0)). This contrasts with the cultivars released over the other three breeding periods, in which the decay of LD was gradual in the range of 3,000 bp (Fig. [5\)](#page-8-0).

Discussion

This study represents a large resequencing effort to characterize the *Lr21* gene in an improved wheat gene pool and the first comprehensive analysis to illustrate the patterns of nucleotide sequence variation at a selected locus influenced by the long-term plant breeding. The analysis revealed 13 SNPs, 4 indels, 10 haplotypes, and substantial genetic changes at the locus. Non-synonymous polymorphic sites and haplotype numbers were increased over the 100 years of wheat breeding. Nucleotide diversity was reduced for the cultivars released before 1993 and increased for those after 1993. Positive selection was observed in HRS wheat cultivars released before 1935. At least two recombination events were inferred in winter and HRS wheat cultivars released before 1994. LD was decreased over time. These findings demonstrate not only the effectiveness of the wheat breeding in the improvement of leaf rust resistance, but also help to understand the genetic impacts of a long-term artificial selection on single loci (Zeng and Cockerham [1990](#page-10-27); Walsh [2005](#page-10-28)).

Direct breeding efforts to incorporate leaf rust resistance genes at the *Lr21* locus into hexaploid wheat were made in 1970s by introgressing the resistant *Ae. tauschii* line into the early dominant cultivar Thatcher (Rowland and Kerber [1974](#page-10-11)). This introgression generated several cultivars with the functional *Lr21* allele with improved resistance to leaf

Fig. 5 The decay of linkage disequilibrium at the *Lr21* locus in 95 wheat cultivars. Five wheat cultivar groups shown are all 95 cultivars, cultivars before 1935, cultivars from 1935 to 1958, cultivars from 1959 to 1993, and cultivars after 1993

rust (Fig. [4](#page-6-1)). Continued incorporation of additional leaf rust resistance genes (e.g., *Lr39*) from diverse germplasm, particularly from CIMMYT and the USA, would also have introduced different $lr21$ alleles (Singh et al. 2004). Consequently, many cultivars released after 1993 displayed more non-synonymous polymorphic sites, more haplotype numbers, and higher nucleotide diversity than the cultivars released before 1994 (Table 3). Also, marginally significant positive selection and decreased LD were also detected in the cultivars released after 1993 (Table [3\)](#page-7-1). Thus, these genetic changes clearly illustrate the effectiveness of the exotic gene introgression in the improvement of leaf rust resistance in hexaploid wheat.

The breeding efforts before 1970 were not directly intended to modify the genes at the *Lr21* locus, but they also generated a substantial impact on the locus. Recombination occurred in winter wheat cultivars released from 1935 to 1958 and in HRS wheat cultivars released from 1959 to 1993. Positive selection was detected in HRS wheat cultivars released before 1935. Nucleotide diversity was gradually reduced for those cultivars released before 1994. These genetic changes may reflect the consequences of hybridization and linkage drag (Hanson [1959](#page-9-15); Robertson [1960;](#page-10-30) Hill and Rasbash [1986\)](#page-9-16) introduced by various breeding efforts before 1970. These breeding efforts may have targeted other genes located on the chromosome 1DS, selected the chromosomal segments where the *Lr21* locus resides through repeated backcrossing, and indirectly introduced different *Lr21* alleles into the breeding lines. Thus, a recombination of different *Lr21* alleles could occur, even without direct selection on the *Lr21* gene.

To understand how breeding efforts indirectly influenced the *Lr21* locus, we surveyed the genes reported on the same chromosome 1DS from the GrainGenes database (<http://wheat.pw.usda.gov/GG2/index.shtml>), and found three storage proteins (triticins, glutenins, gliadins), two isozymes (hexokinase, glucosephosphate isomerase), two stem rust variants (*Sr18*, *Sr33*), and one red glume locus. One aphid (Liu et al. [2001](#page-10-31)) and one powdery mildew (Huang et al. [2000\)](#page-9-17) resistance genes were also reported on 1DS. These reported genes are largely associated with grain quality and stem rust resistance. Thus, it is possible that the breeding efforts before 1970 for the improvement of grain quality and stem rust resistance played an indirect role in influencing the *Lr21* locus. This reasoning matches well with the dominance of Thatcher, with it superior quality and durable stem rust resistance, in Canadian wheat breeding programs from 1940 to 1970. Extensive backcrossing to Thatcher and frequent utilization of its backcross derivatives (McCallum and DePauw [2008\)](#page-10-16) may have selected and/or maintained the genes associated with grain quality and stem rust resistance on the chromosome 1DS,

and consequently influenced indirectly the nearby *Lr21* locus.

Interestingly, no partially functional haplotypes were identified in these assayed cultivars. The haplotype 1 is fully functional, while the other nine haplotypes are considered to be pseudogenes due to sequence truncation and frame shift before the LRR region (Huang et al. [2009](#page-10-13)). A partially functional haplotype would be the derived haplotype 1 with some deletion in the LRR region. It is possible that such a partially functional haplotype exists, given the levels of intragenic recombination observed here and reported by Huang et al. ([2009\)](#page-10-13). Also, the haplotype 2 as represented by Red Fife was the common founder allele in hexaploid wheat. Linking this finding to those reported *Ae. tauschii* alleles (Huang et al. [2009](#page-10-13)) suggests that more than two hybridizations may have been involved with wheat evolution. Thus, tracing the haplotype 2 in other grass plants may shed some light on the evolution of wheat and other cereal crops at the *Lr21* locus.

The characterization of the *Lr21* gene in this set of diverse wheat germplasm is significant for further improvement of leaf rust resistance. More haplotypes were found and four major haplotype groups were named. As only one functional haplotype is identified so far, associating detected haplotypes with the field leaf rust resistance rat-ings (Table [1\)](#page-2-0) should offer a useful means to separate the rust resistance conditioned by genes from different loci. For example, such an association confirmed that the leaf rust resistance in cultivar Pasqua carrying *Lr11, Lr13, Lr14b, Lr30,* and *Lr34* (Dyck [1993](#page-9-18); McCallum et al. [2007\)](#page-10-12) was not conditioned by genes at the *Lr21* locus. The newly developed SCAR marker can be applied to identify the resistant haplotype and four named haplotype groups. Thus, it could supplement the use of the early released diagnostic marker (KSUD14-STS; Huang and Gill [2001](#page-9-7)) not only in wheat, barley, and oat breeding, but also in the inference of *Lr21* gene evolution.

The findings presented here also demonstrate the usefulness of a population-based resequencing analysis of improved crop gene pools in understanding the effectiveness of plant breeding in genetic modification. With more breeding target genes being cloned and the reduced cost of DNA sequencing, this approach will become more practical and feasible. Applications of this approach will help to identify alleles of interest for plant breeding, to reveal detailed genetic changes at breeding target loci in an improved crop gene pool or breeding program, and to understand the genetic impacts of a long-term artificial selection on plant genomes.

Acknowledgments Authors would like to thank Dr. Ron DePauw for his help in wheat class identification; Mr. Dallas Kessler and Mr. Eugene Timmermans for their assistance in accession identification and rust resistance data extraction from GRIN-CA; Ms. Marcie Heggie for her technical assistance in DNA sequencing; and three anonymous reviewers for their constructive comments on an early version of the manuscript.

References

- Aranzana MJ, Kim S, Zhao K et al (2005) Genome-wide association mapping in *Arabidopsis thaliana* identifies previously known genes responsible for variation in flowering time and pathogen resistance. PLoS Genet 1:e60
- Borlaug N (2007) Sixty-two years of fighting hunger: personal recollections. Euphytica 157:287–297
- Chen H, Morrell PL, de la Cruz M, Clegg MT (2008) Nucleotide diversity and linkage disequilibrium in wild avocado (*Persea americana* Mill.). J Hered 99:382–389
- Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feuillet C, Keller B, Jordan MC (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large *psr567* gene family. Plant Mol Biol 65:93–106
- DePauw RM, Boughton GR, Knott DR (1995) Hard red spring wheat. In: Slinkard AE, Knott DR (eds) Harvest of gold: the history of field crop breeding in Canada. Univ of Saskatchewan, SK, pp 5–35
- Donini P, Law JR, Koebner RMD, Reeves JC, Cooke RJ (2000) Temporal trends in the diversity of UK wheat. Theor Appl Genet 100:912–917
- Dyck PL (1993) The inheritance of leaf rust resistance in the wheat cultivar Pasqua. Can J Plant Sci 73:903–906
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797
- Feuillet C, Travella S, Stein N, Albar L, Nubla A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L.) genome. Proc Natl Acad Sci USA 100:15253–15258
- Fu YB (2006) Impact of plant breeding on genetic diversity of agricultural crops: searching for molecular evidence. Plant Genet Resour 4:71–78
- Fu YB (2010) FPTest: a SAS routine for testing differences in allelic count. Mol Ecol Resour 10:389–392
- Fu YB, Somers DJ (2009) Genome-wide reduction of genetic diversity in wheat breeding. Crop Sci 49:161–168
- Fu YB, Peterson GW, Scoles G, Rossnagel B, Schoen DJ, Richards RW (2003) Allelic diversity changes in 96 Canadian oat cultivars released from 1886 to 2001. Crop Sci 43:1989–1995
- Fu YB, Peterson GW, Richards KW, Somers DJ, DePauw RM, Clarke JM (2005) Allelic reduction and genetic shift in the Canadian red hard spring wheat germplasm released from 1886 to 2004. Theor Appl Genet 110:1505–1516
- Fu YB, Peterson GW, Yu J-K, Gao LF, Jia JZ, Richards KW (2006) Impact of plant breeding on genetic diversity of the Canadian hard red spring wheat germplasm as revealed by EST-derived SSR markers. Theor Appl Genet 112:1239–1247
- Hanson WD (1959) Theoretical distribution of the initial linkage block lengths intact in the gametes of a population intermated for *n* generations. Genetics 44:839–846
- Hill WG, Rasbash J (1986) Models of long term artificial selection in finite population. Genet Res 48:41-50
- Huang L, Gill BS (2001) An RGA-like marker detects all known *Lr21* leaf rust resistance gene family members in *Aegilops tauschii* and wheat. Theor Appl Genet 103:1007–1013
- Huang XQ, Hsam SLK, Zeller FJ, Wenzel G, Mohler V (2000) Molecular mapping of the wheat powdery mildew resistance gene *Pm24* and marker validation for molecular breeding. Theor Appl Genet 101:407–414
- Huang L, Brooks SA, Li W, Fellers J, Trick HN, Gill B (2003) Mapbased cloning of leaf rust resistance gene *Lr21* from the large and polyploidy genome of breed wheat. Genetics 164:655–664
- Huang L, Brooks S, Li W, Fellers J, Nelson JC, Gill B (2009) Evolution of new disease specificity at a simple resistance locus in a crop-weed complex: reconstitution of the *Lr21* gene in wheat. Genetics 182:595–602
- Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147–164
- Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103:16666–16671
- Jones H, Leigh FJ, Mackay I, Bower MA, Smith LMJ, Charles MP, Jones G, Jones MK, Brown TA, Powell W (2008) Population based re-sequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. Mol Biol Evol 25:2211–2219
- Kolmer JA (1996) Genetics of resistance to what leaf rust. Ann Rev Phytopathol 34:435–455
- Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B (2009) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451– 1452
- Liu XM, Smith CM, Gill BS, Tolmay V (2001) Microsatellite markers linked to six Russian wheat aphid resistance genes in wheat. Theor Appl Genet 102:504–510
- Malysheva-Otta L, Ganal MW, Law JR, Reeves JC, Röder MS (2007) Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.). Mol Breed 20:309–322
- Marais GF, McCallum B, Marais AS (2008) Wheat leaf rust resistance gene *Lr59* derived from *Aegilops peregrina*. Plant Breed 127:340–345
- McCallum BD, DePauw RM (2008) A review of wheat cultivars grown in the Canadian prairies. Can J Plant Sci 88:649–677
- McCallum BD, Seto-Goh P (2008) Physiologic specialization of *Puccinia triticina* in Canada in 2005. Can J Plant Pathol 30:124–132
- McCallum BD, Fetch T, Chong J (2007) Cereal rust control in Canada. Aust J Agric Res 58:639–647
- McIntosh RA, Wellings CR, Park RF (1995) Wheat rusts: an atlas of resistance genes. Kluwer, Boston
- Morrison JW (1960) Marquis wheat, a triumph of scientific endeavor. Agric Hist 34:182–188
- National Research Council, Committee on Genetic Vulnerability of Major Crops (1972) Genetic Vulnerability of Major Crops. Natl Acad Sci, Washington, DC
- Neatby KW (1942) New varieties of spring wheat resistant to stem rust in the Canadian West, and their genetical background. Emp J Exp Agric 10:245–252
- Peterson RF, Campbell AB, Hannah AE (1948) A diagrammatic scale for estimating rust intensity on leaves and stems of cereals. Can J Res 26:496–500
- Robertson A (1960) A theory of limits in artificial selection. Proc R Soc Lond B153:234–249
- Rowland GG, Kerber ER (1974) Telocentric mapping in hexaploid wheat of genes for leaf rust resistance and other characters derived from *Aegilops squarrosa*. Can J Genet Cytol 16:137–144
- SAS Institute Inc. (2008) The SAS system for windows V9.2. SAS Institute Incorporated, Cary
- Singh S, Franks CD, Huang L, Brown-Guedira GL, Marshall DS, Gill BS, Fritz A (2004) *Lr41*, *Lr39*, and a leaf rust resistance gene from *Aegilops cylindrica* may be allelic and are located on wheat chromosome 2DS. Theor Appl Genet 108:586–591
- Slinkard AE, Knott DR (1995) Harvest of gold: the history of field crop breeding in Canada. Univ of Saskatchewan, SK
- Tajima $F(1983)$ Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
- Topol EJ, Frazer KL (2007) The resequencing imperative. Nat Genet 39:439–440
- Tripp R (1996) Biodiversity and modern crop varieties: sharpening the debate. Agric Hum Values 13:48–63
- Wall JD (1999) Recombination and the power of statistical tests of neutrality. Genet Res 74:65–79
- Walsh B (2005) Population- and quantitative-genetic models of selection limits. Plant Breed Rev 24:177–225
- Zeng Z-B, Cockerham CC (1990) Long-term response to artificial selection with multiple alleles—study by simulation. Theor Popul Biol 37:254–272