Molecular Phylogenetics and Evolution 54 (2010) 10–22

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic relationships and reticulation among Asian *Elymus* (Poaceae) allotetraploids: Analyses of three nuclear gene trees

Roberta J. Mason-Gamer *, Melissa M. Burns [1], Marianna Naum [2]

Department of Biological Sciences, The University of Illinois at Chicago, MC 066, Chicago, IL 60607, USA

ARTICLE INFO

Article history:
Received 19 January 2009
Revised 17 September 2009
Accepted 3 October 2009
Available online 8 October 2009

Keywords:
Elymus
Phylogeny
Polyploidy
Reticulation
Triticeae

ABSTRACT

This phylogenetic study focuses on a subset of the species in *Elymus*—specifically, the endemic Asian tetraploids presumed to combine the **St** genome from *Pseudoroegneria* with the **Y** genome from an unknown donor. The primary goals were to (1) determine whether the **St** and **Y** genomes are derived from phylogenetically distinct donors; (2) identify the closest relative, and potentially the likely donor, of the **Y** genome; and (3) interpret variation among **StStYY** species in terms of multiple origins and/or introgression. The goals were addressed using phylogenetic analyses of sequences from three low-copy nuclear genes: phosphoenolpyruvate carboxylase, β-amylase, and granule-bound starch synthase I. Data sets include 16 **StStYY** individuals representing nine species, along with a broad sample of representatives from most of the monogenomic (i.e., non-allopolyploid) genera in the tribe. To briefly summarize the results: (1) the data clearly support an allopolyploid origin for the Asian tetraploids, involving two distinct donors; (2) the **Y** genome was contributed by a single donor, or multiple closely-related donors; (3) the phylogenetic position of the *Elymus* **Y** genome varies among the three trees and its position is not strongly supported, so the identity of the donor remains a mystery; and (4) conflicts among the gene trees with regard to the **St**-genome sequences suggest introgression involving both *Elymus* and *Pseudoroegneria*.

1. Introduction

Allopolyploid species present numerous challenges to molecular phylogenetic studies. While data from the maternally-inherited chloroplast genome are generally easy to obtain from polyploids, they provide an incomplete picture of polyploid origins. Nuclear data are potentially more informative with respect to the reticulate nature of allopolyploids, but they are more expensive and time-consuming to obtain. Furthermore, data from nuclear markers can be difficult to interpret in groups where ploidy levels are not known at the outset, or when sequences are affected by recombination, copy conversion or loss, or within-genome duplication. In spite of these complications, nuclear sequence data have clarified tangled relationships among polyploids and their progenitors in many plant groups; in recent years, data from single- and low-copy markers have proven especially informative (e.g., Ainouche et al., 2004; Brysting et al., 2007; Emshwiller and Doyle, 2002; Fortune et al., 2008; Ge et al., 1999; Joly et al., 2006; Lihová et al., 2006; Mason-Gamer, 2001, 2008; Petersen et al., 2006; Popp and Oxelman, 2001; Rodríguez and Spooner, 2009; Smedmark et al., 2005; Straub et al., 2006). The present study applies phylogenetic data from three single-copy nuclear loci to a group of Asian allotetraploids in the wheat tribe, Triticeae (Poaceae).

The relationships within the Triticeae have received considerable attention, partly because of the economic importance of the group (which includes wheat, barley, and rye), and partly because the many published Triticeae phylogenetic data sets have failed to converge on a straightforward estimate of the relationships among the non-allopolyploids (Kellogg et al., 1996; Mason-Gamer, 2005; Seberg and Petersen, 2007). The tribe's numerous allopolyploids further complicate phylogenetic analyses, due to their explicitly reticulate origin (Kellogg, 1989; Kellogg et al., 1996). Although the cultivated wheats are the most economically important allopolyploids in the Triticeae, and have thus received considerable attention, the allopolyploid genus *Elymus* L. is in some ways more interesting from an evolutionary standpoint, with its many species, wide natural distribution, morphological variability, and variety of distinct genome combinations (Dewey, 1984; Löve, 1984).

The circumscription of *Elymus* has changed through time, and still varies considerably among treatments in current use (Barkworth, 2000). The practice of defining Triticeae genera by their genomic compositions, as inferred from patterns of chromosome pairing, has been critiqued for several legitimate reasons (Seberg and Petersen, 1998). However, the genomic definition of *Elymus*

* Corresponding author. Address: University of Illinois at Chicago, Department of Biological Sciences, MC 066, 845 W. Taylor St., Chicago, IL 60607, USA. Fax: +1 312 413 2435.
E-mail address: robie@uic.edu (R.J. Mason-Gamer).
[1] Present address: Exilixis, Inc., San Francisco, CA 94083, USA.
[2] Present address: Division of Microbiology, Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD 20740, USA.

1055-7903/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.ympev.2009.10.002

is more reflective of its evolutionary history than are alternative definitions based on morphology, and is the definition followed here. According to its genomic definition (Dewey, 1984; Löve, 1984), *Elymus* comprises about 150 allopolyploid species with at least one set of chromosomes derived from *Pseudoroegneria* (Nevski) Á. Löve (genome designation **St**). In *Elymus*, the **St** genome can be combined with genomes from one or more Triticeae genera, including *Hordeum* L. (genome designation **H**), *Agropyron* Gaertn. (**P**), *Australopyrum* (Tzvelev) Á. Löve (**W**), and an unknown donor (**Y**), in various allopolyploid combinations including **StStHH**, **StStYY**, **StStHHHH**, **StStStStHH**, **StStStStYY**, **StStYYYY**, **StStHHYY**, **StStYYWW**, and **StStYYPP** (Dewey, 1967, 1968, 1970, 1974, 1984; Jensen, 1990, 1993, 1996; Lu et al., 1995; Lu and von Bothmer, 1991, 1993; Salomon, 1993; Salomon and Lu, 1992, 1994a). Other **St**-containing allopolyploids include *Pascopyrum smithii* (Rydb.) Á. Löve, which combines the *Pseudoroegneria* and *Hordeum* genomes with the **Ns** genome of *Psathyrostachys* Nevski in an **StStHHNsNsNsNs** octoploid configuration (Dewey, 1975), and *Thinopyrum* Á. Löve, some species of which are hypothesized to combine the **St** genome with the **E** and/or **J** genomes usually considered characteristic of *Thinopyrum* (e.g., Chen et al., 1998; Liu and Wang, 1993; Zhang et al., 1996). Thus, the **St** genome, probably more than any other in Triticeae, plays an important role in the complex reticulate allopolyploid patterns that characterize the tribe.

The present analyses focus on the **StStYY** *Elymus* tetraploids, which comprise 30–40 species restricted to temperate Asia (Lu and Salomon, 1992). Numerous cytogenetic studies have addressed the genome content of one or more species in this group (Dewey, 1974, 1980a,b; Jensen, 1989, 1990; Jensen and Hatch, 1989; Lu et al., 1990, 1995; Lu and von Bothmer, 1989, 1990a,b, 1991; Sakamoto and Muramatsu, 1966). These studies highlight the lack of pairing between the **St**, **Y**, and/or **H** genomes, suggesting that the **Y** genome donor is unlikely to have been either *Hordeum* (**H**) or *Pseudoroegneria* (**St**). No putative **Y**-genome diploids have been identified, so the origin of the genome remains unknown. Analyses of repetitive DNA markers (Svitashev et al., 1996), and microsatellites and RAPDs (Sun et al., 1997), show a genetic distinction between the genomically-defined **StStYY** and **StStHH** *Elymus* species, and are thus consistent with the cytogenetic studies. More recently, the phylogenetic affinities of the genomes in this group were explored in two studies using DNA sequence data. Both studies included a broad sample of **StStYY** tetraploids (along with other *Elymus* allopolyploids), and a somewhat limited selection of monogenomic (i.e., non-allopolyploid) genera. One of these, based on internal transcribed spacer (ITS) sequences of the nuclear rDNA repeat (Liu et al., 2006) revealed extensive sequence variation within and among sequences from the **StStYY** individuals, but did not uncover any obvious pattern consistent with allopolyploidy (i.e., the sequences did not form distinct clades representing the presumed **St** and **Y** genomes). Based on their results, Liu et al. (2006) concluded that the **Y** genome was probably a derivative of the **St** genome. In contrast, a similar analysis based on sequences of the nuclear gene encoding the b subunit of RNA polymerase II (*RPB2*) showed intra-individual variation consistent with allotetraploidy (Sun et al., 2008): two phylogenetically divergent sequence types were recovered from within individuals, suggesting two distinct donors.

This study further explores the allopolyploid origin of the Asian tetraploid *Elymus* species using data from three low-copy nuclear genes, and a more extensive sample of representatives from most of the monogenomic genera in the tribe. The primary goals were to (1) determine whether the **St** and **Y** genomes are derived from phylogenetically distinct donors; (2) identify the closest relatives (and possible donors) of the genomes—especially the **Y** genome, whose identity has been more elusive; and (3) interpret variation among **StStYY** species in terms of multiple origins and/or introgression. In summary, the data clearly support an allopolyploid origin for the Asian tetraploids, involving donors from distinct clades; the **Y** genome was contributed from a single unknown donor (or multiple closely-related donors); the phylogenetic position of the **Y** genome is variable among trees and lacks support; and the **St** genome sequences have been subject to reticulation, possibly as a result of introgression within and between *Elymus* and *Pseudoroegneria*.

2. Materials and methods

2.1. Samples

The present analyses include 16 tetraploid individuals representing nine presumed **StStYY** tetraploid species (Table 1). This sample includes a broad representation of the geographical range of the **StStYY** group (Lu and Salomon, 1992), and includes representatives of the **StStYY** species groups based on previous analyses of morphology, hybridization, and genome pairing. These include the *E. tibeticus* group (represented here by *E. gmelinii* and *E. pendulinus*; Salomon and Lu, 1994b), the *E. semicostatus* group (*E. abolinii*, *E. nevskii*, and *E. semicostatus*; Salomon and Lu, 1994a,b), and the *E. parviglumis* group (*E. antiquus*, *E. caucasicus* and *E. longearistatus*; Lu and von Bothmer, 1993). (*Elymus ciliaris*, also included here, was not placed within one of these groups.) The present sample is not sufficient to test the groups, nor was that the intent; the proposed groups were used as a guide to ensure a diverse sample.

Three single- or low-copy nuclear genes were amplified, and multiple clones were checked with the goal of obtaining sequences representing both sets of genomes. For each gene, *Elymus* sequences were analyzed along with a reasonably broad sample of the tribe's known genomic diversity, including representatives of 15 monogenomic genera (Table 2). These include the donor of the **St** genome (*Pseudoroegneria*), and all of the genomes known to co-occur with **St** in allopolyploids: *Hordeum* (**H**), *Agropyron* (**P**), *Australopyrum* (**W**), *Psathyrostachys* (**Ns**; this genome is represented by the tetraploid *Leymus racemosus* ssp. *sabulosus* (M. Bieb.) Tzvelev in the pepC data set), and *Thinopyrum* (**J** and/or **E**). No monogenomic **Y**-genome species are known. Additional monogenomic genera included were *Aegilops*, *Crithopsis* (except for pepC), *Dasypyrum*, *Eremopyrum*, *Henrardia* (except for pepC), *Heteranthelium*, *Peridictyon*, *Secale*, *Taeniatherum*, and *Triticum*. While generic definitions within the Triticeae have been extremely variable (reviewed in Barkworth, 2000), this sample represents nearly all of the monogenomic genera accepted in genome-based classifications of the tribe. All three trees were rooted with a representative of *Bromus* L.; Bromeae and Triticeae have been shown to form a single clade, with Bromeae as either sister or paraphyletic to a monophyletic Triticeae (Davis and Soreng, 2007; Grass Phylogeny Working Group, 2001).

2.2. Molecular methods and alignment

All but four of the **StStYY** *Elymus* sequences (Table 1) are newly published here. Nearly all of the sequences from the monogenomic species, with a few exceptions as noted, were previously published in various sources (Table 2). Information about the data and taxa can be found therein, but the primary discussions about the characteristics of each marker and data set are: pepC—(Helfgott and Mason-Gamer, 2004); β-amylase—(Mason-Gamer, 2005); and GBSSI—(Mason-Gamer, 2001; Mason-Gamer et al., 1998). Based on studies of grass genomes in crop species, the three nuclear markers appear to be on three different chromosomes (more below). This is a tentative assumption, based on a small number of