Theor Appl Genet (2015) 128:1865–1879
DOI 10.1007/s00122-015-2553-6

ORIGINAL PAPER

A simple model for pollen-parent fecundity distributions in bee-pollinated forage legume polycrosses

Heathcliffe Riday[1] · Mark A. Smith[2] · Michael D. Peel[3]

Received: 10 April 2015 / Accepted: 5 June 2015 / Published online: 24 June 2015

Abstract

Key message **A simple Weibull distribution based empirical model that predicts pollen-parent fecundity distributions based on polycross size alone has been developed in outbred forage legume species for incorporation into quantitative genetic theory.**

Abstract Random mating or panmixis is a fundamental assumption in quantitative genetic theory. Random mating is sometimes thought to occur in actual fact, although a large body of empirical work shows that this is often not the case in nature. Models have been developed to explain many non-random mating phenomena. This paper measured pollen-parent fecundity distributions among outbred perennial forage legume species [autotetraploid alfalfa (*Medicago sativa* L.), autohexaploid kura clover (*Trifolium ambiguum* M. Bieb.), and diploid red clover (*Trifolium pratense* L.)] in ten polycrosses ranging in size (*N*) from 9 to 94 pollinated with bee pollinators [Bumble Bees (*Bombus impatiens* Cr.) and leafcutter bees (*Megachile rotundata* F.)]. A Weibull distribution best fit the observed pollen-parent fecundity distributions. After standardizing data among the 10 polycrosses, a single Weibull distribution-based model was obtained with an R^2 of 0.978. The model is able to predict pollen-parent fecundity distributions based on polycross size alone. The model predicts that the effective polycross size will be approximately 9 % smaller than under random mating (i.e., $N_e/N \sim 0.91$). The model is simple and can easily be incorporated into other models or simulations requiring a pollen-parent fecundity distribution. Further work is needed to determine how widely applicable the model is.

Communicated by H. H. Geiger.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2553-6) contains supplementary material, which is available to authorized users.

✉ Heathcliffe Riday
heathcliffe.riday@ars.usda.gov

[1] U.S. Dairy Forage Research Center, USDA-ARS, Madison, WI 53706, USA

[2] S&W Seeds, Arlington, WI 53911, USA

[3] Forage and Range Research Lab, USDA-ARS, Logan, UT 84322, USA

Introduction

Random mating is a fundamental assumption underlying most quantitative genetic models (Kempthorne 1969; Falconer and Mackay 1996; Allard 1999). Random mating assumes that in a random mating population every member has an equal chance of producing offspring and that any female gamete is equally likely to be fertilized by any male gamete (Allard 1999). Random mating is mathematically defined as maternal and paternal gamete frequencies contributing equal parentage of progeny in the next generation (i.e., uniform distributions). The concept of random mating was first put forth by Weismann (1883) in his paper "On Heredity" in which he defines the term panmixia, which later became synonymous with the term "random mating". From a philosophical perspective, it is instructive to review Weisman's original definition: "This suspension of the preserving influence of natural selection may be termed Panmixia" (Weismann 1883, p. 90). Philosophically, Weisman's original definition emphasizes panmixis as state of non-selection. With time, as the term random mating replaced panmixis and the random mating assumption became a central dogma in quantitative genetics, the term random mating took on a subtle additional philosophical

emphasis. Rather than only being an abstract state of non-selection, the random mating assumption took on the aspect of being an actual phenomenon in nature. Evidence for this philosophical shift in thinking can be found by conducting literature searches for the phrase "allowed to random mate". Such searchers turn up (usually) plant cultivar registration/development publications where authors using the term "allowed to random mate" are referring to allowing selected genotypes to cross pollinate (usually in isolation) with the breeder providing no selection pressure (e.g., Bartlerr and Butler 1975; Globerson et al. 1987; Rai et al. 1994; Lambert and Chung 1995; Atugwu and Uguru 2012). In other words, breeder/human unassisted/non-interventionary selection became conflated with the quantitative genetic concept of random mating (i.e., panmixis). Allard (1999) warns against this conflation after defining the term random mating: "It can be questioned whether the theoretical form of random mating is ever fulfilled exactly, because some form of selection, natural or human, is likely to intervene…environmentally induced differences…make it improbable that fertilizations are ever entirely random events, either in nature or in cultivation" (Allard 1999, p. 53). This conflation is not by any means universal, but appears particularly confined to applied plant breeding literature, as there is an extensive existent body of literature showing deviations from theoretical random mating in ecological and evolutionary literature (Partridge 1983; Jiang et al. 2013). For plants growing in a natural environment, the assumption that all pollen grains are equally likely to pollinate all other plants has repeatedly been shown to not be the case (Clegg 1980; Allard 1999; Nyquist and Santini 2007; Edelaar and Bolnick 2012).

Clegg (1980) proposes that the utility of the random mating assumption is its use as a null hypothesis in trying to determine significant deviation from theoretical random mating. Examples of studied non-random mating systems with clear violations of panmixis include: assortative mating, self-pollination systems, and sub-clustering of mating individuals within populations due to distance between mating individuals (Clegg 1980; Partridge 1983; Falconer and MacKay 1996 Allard 1999). In plants, the question can be asked: if the possibility of self-pollination is removed, and the non-random mating expectation is low (i.e., insignificant isolation by distance, no expectation of assortative mating), and mating individuals are in a confined mating space, would theoretical random mating be observed? Germane to this question is research looking at lifetime fecundity in plants including closely related work on fecundity hierarchies (Gottlieb 1977; Solbrig and Solbrig 1984; Heywood 1986; Scheiner 1987; Dodd and Silvertown 2000; Herrera and Jovani 2010; Chybicki and Burczyk 2013). Heywood's (1986) work is particularly relevant in which he modeled effective population size reduction due to genetic drift when taking into account the fecundity distribution of the parental generation. He indicates this distribution has two components: one based on differential parent contribution (i.e., genetic, fitness) and one based on random success of different parents (i.e., stochastic). Much of the work on lifetime fecundity and fecundity hierarchies examines effects of plant size on fecundity, particularly in relationship to plants in densely growing situations with intra-species competition. Significant fecundity distribution variation between annual and perennial plants and between species has been observed (Heywood 1986; Dodd and Silvertown 2000; Solbrig and Solbrig 1984; Herrera and Jovani 2010). However, much of this research assumes that fecundity distributions are log-normal (Solbrig and Solbrig 1984; Heywood 1986; Scheiner 1987; Herrera and Jovani 2010; Chybicki and Burczyk 2013).

In studies that have examined inbreeding effects due to fecundity distributions, most often Sewall Wright's (1922) coefficient of inbreeding approach is used in terms of measuring and modeling expected allele frequencies and concomitant estimates of homo- and heterozygosity and various inbreeding measures. Few studies have used a coancestry approach (Malécot 1948; Kempthorne 1969); this is likely because research often examines populations in the wild. Under open-pollination conditions in nature, it is often difficult to determine relationships between all mating individuals or even to sample all possible mating individuals. Therefore, researchers will sample individual loci within genotypes in a natural setting and using allele-frequencies and Hardy–Weinberg assumptions will pursue Sewall Wrights' coefficient of inbreeding approach. However, in a plant breeding setting, all mating individuals are usually known; in such situations, it is possible and easier to sample genotypes rather than loci within genotypes and to estimate frequencies of various relationship classes within the populations using a coancestry approach. Chybicki and Burczyk (2013) did follow a coancestry-type approach in their oak tree mating study with their models taking into account that the fecundity distribution would not necessarily be uniform (i.e., panmixis or random mating). However, they also assumed that the fecundity distribution would be log-normal presumably based on previous research in the area of hierarchical fecundity.

In this study we wanted to take an empirical approach to examine bee-mediated mating in perennial forage legume species, particularly breeding polycrosses (i.e., so-called "random" mating populations) with <100 mating genotypes under uniformly moderate to minor intra-specific competition between mating individuals and where the total pollination area is restricted to a very limited space. We wanted to determine if theoretical random mating (panmixis) was observed in such situations. If panmixis was not observed, we wanted to empirically determine a model that would reasonably predict the observed pollen-parent

Table 1 Ten polycrosses: name, size (*N*), plant species, progeny number sampled, initial number of parents planted, seed-parents sampled, pollinator used, polycross physical size, polycross isolation, and polycross location

Polycross name	Size (*N*)	Species	Progeny sampled[a]	Initial parents planted[b]	Seed-parents sampled	Pollinator bee	Physical size	Polycross isolation	Polycross location
Pioneer	9	Alfalfa	763	9	6	Bumble	1.22 m × 1.22 m[c]	Indoor room	Arlington, WI
Alforex	15	Alfalfa	536	16	15	Leafcutter	2.74 m × 18.29 m	Screened cage	Woodward, CA
LP08	22	Red Clover	321	30	8	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
KU09B	26	Kura Clover	656	27	26	Leafcutter & Natural	9.14 m × 18.29 m	Open isolation	Millville, UT
Vis09	26	Red Clover	680	33	11	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
C328WS	34	Red Clover	604	40	12	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
Yld09	74	Red Clover	513	96	14	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
C584Y	93	Red Clover	1103	96	27	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
WI21	93	Red Clover	1704	96	31	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI
C276	94	Red Clover	1043	96	27	Bumble	1.83 m × 3.66 m	Screened cage	Prairie du Sac, WI

[a] Numbers include only successfully paternity-tested progeny; in alfalfa polycrosses numbers exclude selfed progeny

[b] Of parents established in a given polycross, not all plants survived to reproduce. Therefore, polycross size (*N*) was determined based on paternal parents detected during paternity testing of progeny and seed-parents which produced seed

[c] Dimensions include immediate size of polycross; actual indoor room size was larger

fecundity distributions. We sampled progeny produced from 10 polycrosses of varying size (9–94 pollen-parents). Each polycross was conducted in some form of isolation with all potential parents known. Polycrosses were sampled from three forage legume species: alfalfa (*Medicago sativa* L., $2n = 4x = 32$), kura clover (*Trifolium ambiguum* M. Bieb., $2n = 6x = 48$), and red clover (*Trifolium pratense* L., $2n = 2x = 14$). Pollinators used in various polycrosses included: bumble bees (*Bombus impatiens* Cr.), leafcutter bees (*Megachile rotundata* F.), and naturally present pollinators. Our objectives were to (1) determine pollen-parent fecundity distributions for the sampled polycrosses and (2) determine a model that depends only on the number of possible pollen-parents in the polycross to predict the pollen-parent fecundity distribution.

Materials and methods

Polycrosses

Progeny from ten polycrosses were sampled (Tables 1, 2). Two of the polycrosses contained alfalfa, one contained kura clover, and seven contained red clover (Table 1). The number of parents given for each polycross included only parents who were detected as fathers (pollen-parents) among the progeny (Table 1). The actual number of parents placed in the polycross in some cases was greater than the number given for each polycross (Table 1). This was particularly true for many of the red clover polycrosses which contained plants that had survived multiple years in selection nurseries and were subsequently dug up and transplanted together in crossing blocks. A good number of these unreplicated dug and subsequently transplanted red clover plants died prior to pollination (Table 2). Therefore for analysis purposes, we based polycross size on the number of plants that were detected as pollen-parents in the progeny. Details of each polycross follow.

The "Pioneer" alfalfa polycross was a nine parent synthetic of which seed from six plants were sampled (Table 1). The polycross was conducted during Winter 2011/2012 in an indoor room at the DuPont-Pioneer alfalfa breeding station in Arlington, Wisconsin (43°20′18″N, 89°24′14″W). Seven clonal copies of the nine parents were grown in 12.5 cm diameter round pots (Table 2). Plants were staked up to keep plant foliage separated. The 63 plants were completely randomized prior to each pollination. The 63 plants were divided into sets of eight and

Table 2 Ten polycrosses sampled with a given polycross': name, parental plant source, if parental plants were clonally replicated, if seed-parents were uniformly sampled, and source of progeny DNA for analysis

Polycross name	Parent plant source	Parent clonal replicated?	Seed parent uniformly sampled?[b]	Progeny DNA sample sources
Pioneer	Stem cuttings	Yes	$P > 0.25$	Greenhouse seedlings (50 % pre- & 50 % post-disease screen)
Alforex	Stem cuttings	Yes[a]	$P < 0.0001$	Field established transplants
LP08	Nursery dug	No	$P < 0.0001$	Field established transplants
KU09B	Clonal copies	Yes	$P > 0.5$	Directly from seed
Vis09	Nursery dug	No	$P > 0.5$	Field established transplants
C328WS	Field dug	No	$P < 0.0001$	Field established transplants
Yld09	Nursery dug	No	$P < 0.01$	Field established transplants
C584Y	From seed	No	$P < 0.0001$	Field established transplants
WI21	From seed	No	$P < 0.0001$	Field established transplants
C276	From seed	No	$P < 0.0001$	Field established transplants

[a] Parental clonal replication was accomplished; however clonal-copies were planted next to each other in a clonal row

[b] χ^2-test P values indicating if progeny number sampled from seed-parent in each polycross deviates from a uniform distribution

placed in trays (2 × 4 plants) with pots within trays touching each other. The eight trays were separated by approximately 5 cm and placed in the pollination room on tables. A Class A bumble bee hive (approx. 150–200 bees) (Koppert biological, Howell, MI) was placed in the room and the bumble bees were allowed to pollinate the crossing block for 20–24 h (Table 1). The 63 plants were pollinated four days with different randomizations. After pollination, seed was allowed to ripen and seed from clonal copies was bulked to form nine lots of halfsib seed, one for each of the nine parents. Progeny from six of the nine parents were sampled. Paternity testing and associated DNA marker techniques used are described in the next section. Paternity was determined on 763 outcross progeny from this polycross (Table 1). Self-pollination occurs at low frequencies in bee-pollinated alfalfa (Riday et al. 2015). Progeny that were identified as self-pollinations were excluded from analysis in this paper and did not count towards the 763 progeny used from this polycross (The self-pollination rate observed in this polycross was 7.4 %). Of the 763 progeny, 394 of the progeny were sampled directly from seedlings grown from seed, while 369 of the progeny were sampled from progeny that had successfully survived a disease resistance screening.

The "Alforex" polycross was a 15 parent synthetic of which seed from all 15 parents were sampled (Tables 1, 2). The polycross was conducted at the Alforex breeding station in Woodland, California (38°37′05″N, 121°47′55″W) in a Capay silt clay (fine, montmorillonitic, thermic Typic Chromoxererts). The details of this polycross and associated paternity testing and DNA marker techniques are describe in detail in Riday et al. (2013). From this polycross, 536 outcross progeny with identified paternity were used. As with the other alfalfa polycross, self-pollinations were not numbered among the 536 progeny used.

The "LP08" polycross was a 22 parent red clover polycross from which 8 seed parents were sampled (Tables 1, 2). Parent plants for these polycrosses were dug in 2008 out of a 2004-established mixed red clover/tall-fescue (*Festuca arundinacea* Schreb.) grazing study at Lancaster, Wisconsin (Riday et al. 2007). Parent plants were dug the first week in May of 2008 and transplanted the next day at the U.S. Dairy Forage Research Farm in Prairie du Sac, Wisconsin (43°20′52″N, 89°45′24″W) in a Richwood silt loam soil (Fine-silty, mixed, superactive, mesic Typic Argiudolls). Plants in the crossing block were planted in a 3 × 10 plant grid with 45 cm between the 3 rows and 30 cm between the 10 plants within the 3 rows. Plants were staked-up, mulched, enclosed, and pollinated using techniques described in Riday and Krohn (2010). Out of 30 plants, 22 survived to participate in pollination. The crossing block was enclosed in mid-June of 2008 and pollination was allowed to proceed through early August 2008 when pollinators were removed and halfsib seed was harvested from each plant separately. 321 progeny sampled from 8 seed parents from LP08 were genotyped and paternity tested (details provided in next section).

The "KU09B" kura clover polycross was a 26 parent synthetic of which seed from all 26 parents were sampled (Tables 1, 2). The polycross was conducted at Utah State University's Evans Research Farm located in Millville, Utah (41°41′36″N, 111°49′55″W) in a Nibley silty clay loam (fine, mixed, mesic Aquic Argiustolls). Parent plants for this polycross were selected from a breeding population established to produce breeder seed. Final plant selections were made in the fall of 2009 and cuttings were taken from

27 plants and clonally propagated in Ray Leach Cone-tainers (Stewe and Sons, Corvallis, OR) in the greenhouse during the 2009/2010 winter. Clones were then transplanted to the field on May 12, 2010 in rows one meter apart with one meter between plants within a row in a 9 × 18 plant grid. Six clones of each original parental plant were included in a randomized complete block design where each clonal copy represented one replication. Plants within replicates were arranged in one 1 × 18 plant plus 1 × 9 plant arrangement. Originally the polycross included 6 clonal replicates of the 27 parent plants; however, the six clonal copies of one of the 27 parental plants were culled, bringing the number of plants participating in pollination to 6 clones × 26 parental plants. Seed was produced during the 2011 growing season. Kura clover was not common in the pollination area and no known plants were found within a 400 meter radius of the polycross. Commercially purchased leafcutter bees were stationed approximately 15 m north of the crossing block in a shelter with nesting boards. The bees were stationed in front of a Utah sweetvetch (*Hedysarum boreale* Nutt.) seed increase which started flowering well before the kura clover started flowering. The kura clover started flowering during the last week of June. Wild pollinators were abundant in the area; these and honey bees (*Apis mellifera* L.) were observed visiting the plants during flowering. From this polycross 656 progeny were genotyped and paternity tested as described in the next sections.

The "Vis09" and "Yld09" polycross was 26 parent and 74 parent red clover polycrosses from which 11 and 14 seed parents were sampled respectively (Tables 1, 2). Parent plants for these polycrosses were dug in autumn 2008 out of multiple red clover breeding nurseries established during the Spring of 2005 and evaluated through Autumn of 2008. Plants overwintered in the greenhouse and were transplanted in early April of 2009 at the U.S. Dairy Forage Research Center Farm in Prairie du Sac, Wisconsin (43°20′52″N, 89°45′24″W) in a Richwood silt loam soil (Fine-silty, mixed, superactive, mesic Typic Argiudolls). The Vis09 parent plants were planted in a 3 × 11 plant grid with 45 cm between plants in the 3 rows and 30 cm between the 11 plants within the 3 rows. The Yld09 parent plants were planted in a 6 × 16 plant grid with 20 cm between rows 1 and 2, rows 3 and 4, and rows 5 and 6 among the 6 rows and 30 cm between rows 2 and 3, and rows 4 and 5. Within the 6 rows of the Yld09 polycross the 16 plants were spaced 20 cm apart. Plants were staked-up, mulched, enclosed, and pollinated using techniques and pollinators described in Riday and Krohn (2010). Out of 33 Vis09 parent plants, 26 survived to participate in pollination while 74 out of 96 Yld09 parent plants survived to participate in pollination. The crossing blocks were enclosed in mid-June of 2009 and pollination was allowed to proceed through early August 2009 when pollinators were removed and halfsib seed was harvested from each plant separately. A total of 680 Vis09 progeny sampled from 11 seed parents and 513 Yld09 progeny sampled from 14 seed parents were genotyped and paternity tested (details provided in next sections).

The "C328WS" polycross was a 34 parent synthetic from which progeny of 12 seed-parents were sampled (Tables 1, 2). Plants were dug during the first week of May 2009 out of an autumn 2008 drill-seeded red clover variety trial planted at West Salem, Wisconsin. Plants were transplanted the next day at the U.S. Dairy Forage Research Center Farm at Prairie du Sac, Wisconsin (43°20′52″N, 89°45′24″W) in a Richwood silt loam soil (Fine-silty, mixed, superactive, mesic Typic Argiudolls). The C328WS parent plants were planted in a 4 × 10 plant grid with 30 cm between plants in the 4 rows and 30 cm between the 10 plants within the 4 rows. Plants were staked-up, mulched, enclosed, and pollinated using techniques and pollinators described in Riday and Krohn (2010). A total of 34 out of 40 parent plants survived to participate in pollination. The crossing blocks were enclosed in mid-June of 2009 and pollination was allowed to proceed through early August 2009 when pollinators were removed and halfsib seed was harvested from each plant separately. 604 C328WS progeny were sampled from 12 seed parents and were genotyped and paternity tested (details provided in "DNA extraction, PCR, PCR product evaluation, and paternity testing" section).

The "C276", "C584Y07", and "WI21" polycrosses were 94, 93, and 93 parent red clover polycrosses respectively (Tables 1, 2). Seed from 27, 27, and 31 seed-parents were sampled for C276, C584Y07, and WI21 respectively (Table 1). The details of this polycross are described in Riday (2011). From these polycrosses a total of 1043, 1103, and 1704 progeny were sampled from C276, C584Y07, and WI21 respectively (Table 1). The number of seed-parents and progeny analyzed are greater than the numbers reported in Riday (2011); this is because additional seed parents and progeny from these polycrosses have been analyzed for paternity since the publication of this previous study. The newer progeny analyzed for these polycrosses were genotyped using a newer set of DNA markers described below.

DNA extraction, PCR, PCR product evaluation, and paternity testing

Parents and progeny for polycrosses: Pioneer, Alforex, LP08, Vis09, C328WS, and Yld09 were extracted from tissue using methodologies described in Riday et al. (2013). The parents and majority of progeny of C584Y, WI21, and C276 were extracted as described in Riday (2011); additional progeny DNA samples analyzed since the 2011 paper were extracted using methodologies described in

Riday et al. (2013). Finally, parent DNA from the KU09B polycross was extracted as described in Riday et al. (2013), while progeny DNA extracted directly from seed was accomplished as described in Riday et al. (2015).

PCR methodology and PCR product evaluation for the Alforex polycross are described in Riday et al. (2013). PCR methodology and PCR product evaluation for the parents and majority of C584Y, WI21, and C276 progeny are described in Riday (2011). Progeny and parent samples from the LP08, Vis09, C328WS, and Yld09 polycrosses and the progeny samples from the C584Y, WI21, and C276 polycrosses analyzed after publication of Riday (2011), along with the polycross parents were subjected to a single multiplex PCR reaction per sample performed in a 6 µl reaction volume using 15 primer pairs (Table 3; primer sequences available in Sato et al. 2005; Isobe et al. 2009; and online at http://marker.kazusa.or.jp/Red_clover/). For the Pioneer polycross a single 6 µl reaction volume using 16 primer pairs was developed (Table 3; primer sequences available in Sledge et al. 2005). Finally for KU09B we developed a single 6 µl reaction volume using 17 primer pairs (Table 3; primer sequences available in Sato et al. 2005; Isobe et al. 2009; and online at http://marker.kazusa.or.jp/Red_clover/). During multiplex development, each primer pair was amplified individually to determine its pattern and verify its products when multiplexed. Criteria for primer pair inclusion in multiplexes were: number of alleles observed per SSR locus, size of amplification product, and ability to amplify as expected when included in the multiplex. Trial and error was used to determine amount of each primer pair utilized in the multiplex in order to adjusted PCR amplicon production amounts to be more uniform across all primer pairs. The forward primer in each primer pair was fluorescently labeled with either HEX (green), 6-FAM (blue), TAMARA (yellow) (Eurofins MWG Operon, Huntsville, AL), or CAL Fluor Red 610 (red) (Biosearch Technologies Inc., Novato, CA). Each PCR reaction contained: 3 µl JumpStart REDTaq ReadyMix (Sigma-Aldrich, St. Louis, MO), 0.05–0.63 µM of each primer (see Table 1 for specific amount per primer pair and fluorescent label used for that primer), 0.63 mM MgCl2, 0.50 M betaine (Sigma-Aldrich, St. Louis, MO), and approximately 10–80 ng template DNA. The primers and DNA were dissolved in TE (pH 8.0), which contributed 1.43 µl of TE (pH 8.0) to the 6 µl reaction. Plates were sealed with polymer seal mats to prevent evaporation during PCR reactions (BioExpress, Kaysville, UT). Thermal cycling was carried out on a DNA Engine Dyad (Bio-Rad Laboratories Inc., Hercules, CA) as follows: 95 °C for 1 min; 45 cycles of 95 °C for 20 s, 50 °C for 2 min, and 72 °C for 1 min; 55 °C for 1 min; 72 °C for 10 min; and a final step of 4 °C for 1 min. Samples were prepared for fragment size determination by combining 1 µl of PCR reaction with 9.81 µl deionized formamide (Life Technologies Inc., Carlsbad, CA) and 0.19 µl size standard. In 4-dye reactions ROX size standard (red fluorescent color) was used while in five color reactions a custom RadiantDy 632 (orange fluorescent color) size standard (75, 100, 140, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, and 500 bp) (BioVentures Inc., Murfreesboro, TN) was used. Fragment sizes were determined on an ABI Prism 3730xl Genetic Analyzer (Applied Biosystems Inc., Foster City, CA). Raw image files were imported into the GeneMarker 1.91 software program (SoftGenetics LLC, State College, PA), and for 5-dye runs a manual color matrix adjustment was made to reduce pull-up between colors as presented in Riday et al. (2015). GeneMarker was then used to interpret the electropherograms and assign alleles to experimental data.

For the seven diploid red clover polycrosses the Cervus paternity testing software was used (Kalinowski et al. 2007). For the two autotetraploid alfalfa polycrosses an exclusion-based paternity testing programmed using the SAS statistical software package (SAS 2012) was used; the calculation methodologies along with SAS code are available in Riday et al. (2013). For KU09B, an autohexaploid kura clover polycross, the exclusion-based paternity testing methodology and SAS code from Riday et al. (2013) was modified to create a specific program amenable to paternity testing in hexaploid species. The SAS code is provided as an accompanying supplemental file. To our knowledge this may be the first paternity testing program written specifically for autohexaploid marker data.

Statistical analysis

Based on paternity testing results the parental contribution frequency to the progeny generation for each pollen-parent in the polycross (Pf_i) was estimated using Proc Freq feature of the SAS software package (SAS 2012). We made a general attempt to sample equally from seed-parents in a polycross (i.e., achieve a uniform distribution). However, because many of the polycrosses included in this study are part of an active breeding program, in some of the polycrosses one or two seed-parents were more represented than the rest. A χ^2-test using the Proc Freq feature of SAS software package was conducted to determine for which polycrosses progeny sampling from seed-parents deviated from uniformity (Table 2).

The Proc Freq feature of the SAS software package was used to determine if Pf_i values in each of the 10 polycrosses were uniformly distributed (i.e., χ^2-test for equal contribution) (Table 4). The Proc Univariate histogram option of the SAS software package was used to determine if Pf_i values in each of the 10 polycrosses deviated from normal, log-normal, Weibull (1951), or gamma distributions using the Anderson–Darling goodness of fit test (Table 4).

Table 3 Three SSR multiplexed PCR reactions used for red clover (*Trifolium pratense* L.), kura clover (*Trifolium ambiguum* M. Bieb.) and alfalfa (*Medicago sativa* L.) genotyping

SSR primer pair	Accession name	Fluorescent label	Final conc. (μM)	Alleles	Fragment sizes
Red Clover (15SSR Multiplex; ROX size standard)					
RCS1683	bb932926	5-HEX	0.21	6	160–175
RCS1737	bb931457	TAMRA	0.25	7	80–98
RCS2987	bb938238	TAMRA	0.27	16	326–377
RCS3805	bb903332	5-HEX	0.19	15	250–277
RCS3858	bb903773	TAMRA	0.23	11	101–116
RCS3906	bb904089	TAMRA	0.21	11	281–323
RCS4501	bb908062	5-HEX	0.27	10	214–226
RCS4854	de237288	5-HEX	0.08	8	93–114
RCS4956	bb910490	6-FAM	0.25	10	262–294
RCS5615	de238880	6-FAM	0.08	7	152–172
RCS5708	de239036	5-HEX	0.21	19	116–153
RCS5823	de240320	6-FAM	0.21	13	182–216
RCS7110	bb926781	6-FAM	0.12	10	114–128
RCS810	de215917	TAMRA	0.21	42	150–232
RCS884	de216538	5-HEX	0.14	8	189–213
††15Plex-Sp	–	6-FAM	–	8	341–350
Kura Clover (17SSR Multiplex; Custom RadiantDy 632 size standard)					
RCS1597	bb930638	6-FAM	0.21	6	90–119
RCS1864	bb932103	6-FAM	0.21	8	133–157
RCS7219	bb928256	6-FAM	0.21	4	248–263
RCS0884	de216538	5-HEX	0.21	4	221–230
RCS1401	de219880	TAMRA	0.21	5	187–206
RCS1724	bb931291	CAL Fluor Red 610	0.21	6	324–336
RCS2826	bb936332	6-FAM	0.21	5	376–386
RCS4164	bb905847	CAL Fluor Red 610	0.21	5	124–136
RCS4864	de237445	TAMRA	0.21	5	103–115
RCS6686	bb921826	5-HEX	0.21	8	153–175
RCS6986	bb924887	5-HEX	0.63	7	255–288
WCS1012	fy465136	6-FAM	0.21	11	169–213
WCS1748	fy465580	6-FAM	0.21	3	349–354
WCS205	fy466473	6-FAM	0.21	8	293–328
WCS327	fy469251	TAMRA	0.21	10	227–252
WCS762	fy462039	CAL Fluor Red 610	0.21	11	278–307
WCS95	fy467074	TAMRA	0.21	6	140–154
Alfalfa (16SSR Multiplex; Custom RadiantDy 632 size standard)					
AL45	al367160	CAL Fluor Red 610	0.21	6	244–299
AW01	aw559402	CAL Fluor Red 610	0.18	8	163–187
†AW146[a]	aw125937	5-HEX	0.25	2	231–232
AW146[b]	aw125937	5-HEX	0.25	6	265–278
AW235	aw685316	5-HEX	0.05	6	186–199
§AW379	aw776398	TAMRA	0.50	12	139–164
BE119	be322616	5-HEX	0.08	5	140–154
BE323955	be323955	6-FAM	0.13	12	221–251
BE78	be204181	HEX	0.21	14	237–261
BF220	bf646168	TAMRA	0.27	10	275–297
BF85	bf647240	TAMRA	0.21	17	232–262
BG230	bg586805	TAMRA	0.21	9	184–214
BG275	bg646431	6-FAM	0.21	7	136–152
BG280	bg647735	CAL Fluor Red 610	0.10	8	128–146

Table 3 continued

SSR primer pair	Accession name	Fluorescent label	Final conc. (μM)	Alleles	Fragment sizes
BI28	bi309553	6-FAM	0.15	13	156–191
BI54	bi270560	6-FAM	0.19	11	324–351
Mt2A03	Ac125476	6-FAM	‡	7	257–276

† AW146 primer pairs produced two separate amplification products designated AW146[a] and AW146[b]

‡ Asymmetric primer amounts added to obtain better amplification of all multiplex primers when combined (forward 0.21 μM, reverse 0.05 μM)

§ 5 alleles per genotype were often observed at this locus. Each allele at the locus was treated as a separate locus

¶ Number of alleles observed and allele size range represents results from all genotypes in our breeding programs that were ever evaluated using these SSR including genotypes not utilized in this study

†† Unknown polymorphic primer cross amplification product (i.e., adventitious marker) between 2 and 30 primers used

Table 4 Goodness of fit test (χ^2-test for uniform distribution and Anderson–Darling for normal, log-normal, Weibull, and gamma distributions) of actual pollen-parent contribution frequency to progeny parentage (Pf_i) of ten polycrosses and all Pf_i pooled compared to expected: uniform, normal, log-normal, Weibull, and gamma distribution frequencies

Polycross name	No. of pollen-parents	Uniform distribution	Normal distribution	Log-normal distribution	Weibull distribution	Gamma distribution
Pioneer	9	$P < 0.0001$	$P > 0.25$	$P > 0.25$	$P > 0.25$	–[a]
Alforex	15	$P < 0.0001$	$P > 0.25$	$P = 0.014$	$P = 0.069$	$P = 0.059$
LP08	22	$P < 0.0001$	$P = 0.187$	$P = 0.118$	$P > 0.25$	$P > 0.25$
KU09B	26	$P < 0.0001$	$P > 0.25$	$P > 0.25$	$P > 0.25$	$P > 0.25$
Vis09	26	$P < 0.0001$	$P = 0.239$	$P = 0.025$	$P > 0.25$	$P > 0.25$
C328WS	34	$P < 0.0001$	$P < 0.005$	$P > 0.25$	$P > 0.25$	$P > 0.25$
Yld09	74	$P < 0.0001$	$P < 0.005$	$P = 0.008$	$P = 0.026$	$P = 0.032$
C584Y	93	$P < 0.0001$	$P = 0.039$	$P < 0.005$	$P = 0.17$	$P = 0.019$
WI21	93	$P < 0.0001$	$P < 0.005$	$P = 0.007$	$P > 0.25$	$P > 0.25$
C276	94	$P < 0.0001$	$P < 0.005$	$P < 0.005$	$P > 0.25$	$P > 0.25$
Pooled Pf_i	Standardized[b]	$P < 0.0001$[c]	$P < 0.005$	$P < 0.005$	$P = 0.208$	$P = 0.004$

[a] SAS software could not fit a gamma distribution because there were less than 10 observations for this polycross

[b] Pooling all observed Pf_i frequencies from all 10 polycrosses; with Pf_i form each polycross scaled by $\frac{1}{N}$ (i.e., $\frac{Pf_i - \frac{1}{N}}{\frac{1}{N}}$)

[c] Probability of the pooled Pf_i deviating from a uniform distribution was determined by the slope of linear regression ($\frac{Pf_i - \frac{1}{N}}{\frac{1}{N}} = \beta \frac{Pp_i}{N+1} + \varepsilon$) being significantly different from zero

Based on the results of the goodness of fit tests, both the gamma and Weibull distributions appeared to fit the data best. We decided to do further modeling work with the Weibull distribution since its associated functions and mathematics were easier to work with than the gamma distribution. Specifically, we used the Weibull inverse cumulative distribution function (Weibull ICDF), also known as the inverse survival function (Forbes et al. 2011), to empirically model the expected pollen-parent contribution frequency to progeny parentage. The form of Weibull ICDF used was:

$$y = \text{int} + k(-\ln(1-x))^{\frac{1}{\lambda}} \quad (1)$$

where, y is dependent variable, int is the Intercept, k is the shape parameter, x is the independent variable, λ is the scale parameter.

We set the dependent variable in our model (y) as equal to Pf_i. We set the independent variable (x) to equal the array of individual pollen-parents. Since pollen-parent is a categorical variable we first transform it into a quantitative variable.

$$x = \frac{Pp_i}{N+1} \quad (2)$$

where, Pp_i is the ith pollen-parent sequence position in array of all possible pollen-parents (N) ordered from least frequent pollen-parent of progeny in a population (Pf_i, *Minimum*) to most frequent pollen-parent of progeny in a population (Pf_i, *Maximum*). N is the Total number of possible pollen-parents in polycross.

This was done by ordering pollen-parents from least frequent Pf_i to most frequent Pf_i with the new quantitative

variable being (Pp_i) or the pollen-parent sequence position in array of all possible pollen-parents from least frequent to most frequent pollen-parent Pf_i. In order to make Pp_i polycross size neutral, Pp_i was divided by the total number of possible pollen-parents (N) in the polycross plus one. Based on this transformation, no matter how big the polycross, the dependent variable x would be an essentially continuous variable with a value between 0 and 1.

We were not satisfied with the scaling of the independent variable (y) (i.e., Pf_i) since its scale was dependent on polycross size. In order to make y polycross size scale independent, we transformed Pf_i to be a function of the expected panmixis polycross value (i.e., uniform distribution or $\frac{1}{N}$).

$$y = \frac{Pf_i - \frac{1}{N}}{\frac{1}{N}} \tag{3}$$

where, Pf_i is the Frequency of progeny having the ith pollen-parent. N is the Total number of possible pollen parents.

This was done by subtracting the expected panmixis value for the polycross (i.e., $\frac{1}{N}$) from Pf_i and dividing this deviation from panmixis again by the expected panmixis value for the polycross (i.e., $\frac{1}{N}$). Combining Eqs. (1), (2), and (3) yielded the following equation.

$$\frac{Pf_i - \frac{1}{N}}{\frac{1}{N}} = \text{int} + k\left(-\ln\left(1 - \frac{Pp_i}{N+1}\right)\right)^{\frac{1}{\lambda}} \tag{4}$$

The dependent variable y in this equation can have a value from negative one to infinity. In the case where the y equals zero, it indicates that the expected panmixis frequency was observed for that particular pollen-parent. The minimum observable value for y is negative one since the minimum possible value of Pf_i is zero and when Pf_i is zero than the remaining negative expected panmixis frequency divided by the positive expected panmixis frequency will always be negative, no matter how large the polycross size. We therefore fixed the intercept of the model at negative one.

$$\frac{Pf_i - \frac{1}{N}}{\frac{1}{N}} = -1 + k\left(-\ln\left(1 - \frac{Pp_i}{N+1}\right)\right)^{\frac{1}{\lambda}} \tag{5}$$

Using model (5) we were able to compare observed Pf_i values from all ten polycrosses in a standardized space on the same scale. Fitting non-linear model (5) and estimating parameters k and λ was done using Proc Nlmixed feature of the SAS software package; this allowed us to weight each polycross equally while fitting the model, despite the fact that each polycross had a different number of observations in it. Specific Proc Nlmixed SAS code used is presented in "Appendix 1". The 10 polycrosses were treated as a random effect with two separate polycross error terms associated with the k and λ parameters. The covariance between these two error terms was set to zero (i.e., a compound symmetry covariance structure).

Results

For this study paternity (i.e., pollen-parentage) of 7923 progeny with known maternity (i.e., seed-parentage) from ten forage legume species polycrosses was successfully determined (Table 1). The pollen-parent contribution frequency to progeny parentage (Pf_i) was estimated for each polycross. In every single polycross the observed Pf_i values clearly deviated from a uniform distribution ($P < 0.0001$); or, in other words, significantly deviated from panmixis or random mating (Table 4). Since a uniform distribution was not observed and we thought there may likely be a stochastic component to Pf_i values, we expected that perhaps the distribution would shift from a uniform to a normal distribution. Indeed we observed a better Pf_i value distribution fit for a normal distribution compared to a uniform distribution (Table 4). However in five out of the ten polycrosses, Pf_i values significantly deviated from a normal distribution. Furthermore, it was the five polycrosses with the fewest number of pollen-parents (i.e., smallest sample size) whose Pf_i values did not deviate from a normal distribution. The reason a normal distribution appeared not to fit well was that the observed distributions were right skewed (Fig. 1). Right skewed distributions were not totally unexpected based on previous research (Solbrig and Solbrig 1984; Scheiner 1987; Herrera and Jovani 2010; Chybicki and Burczyk 2013). We therefore fitted a log-normal distribution (Table 4), which was only marginally better than the normal distribution. Although it was right skewed, it appeared too "extreme" (i.e., more expected low fecundity individuals were predicted than were actually observed). Finally Weibull and gamma distributions, two other right skewed distributions, were fitted. Both the Weibull and gamma distributions generally showed good fits to individual polycrosses (Table 4). The Weibull distribution showed the best fit for the pooled Pf_i values (Table 4; Fig. 1). We therefore used the Weibull distribution for more detailed analysis; the fact that mathematically the Weibull distribution and its various derived functions are easier to use than the gamma distribution was a motivating factor in our decision as well.

One intent of this study was to graphically view Pf_i values for polycrosses of any size in the same graphical space on the same scale. As described in the materials and methods, we developed a formula (5) to do this. We were intrigued at the consistency of the pollen-parent fecundity distributions across all evaluated polycrosses (Fig. 2).

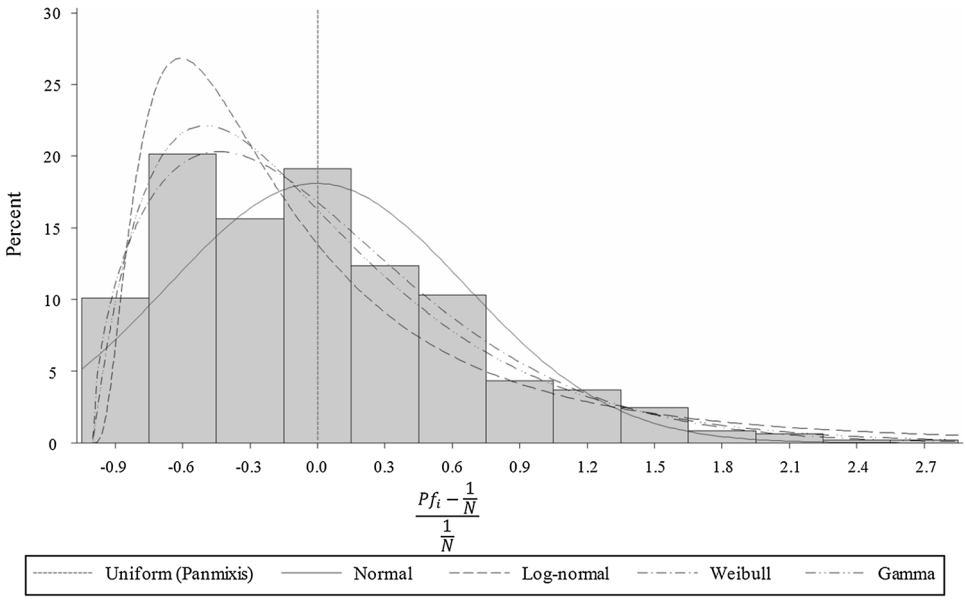

Fig. 1 Histogram of pooled panmixis standardized pollen-parent contribution to progeny frequencies for 10 bee-pollinated forage legume polycrosses ranging in size from 9 to 94 pollen-parents. Fitted distributions include: uniform (i.e., panmixis), normal, log-normal, Weibull, and gamma

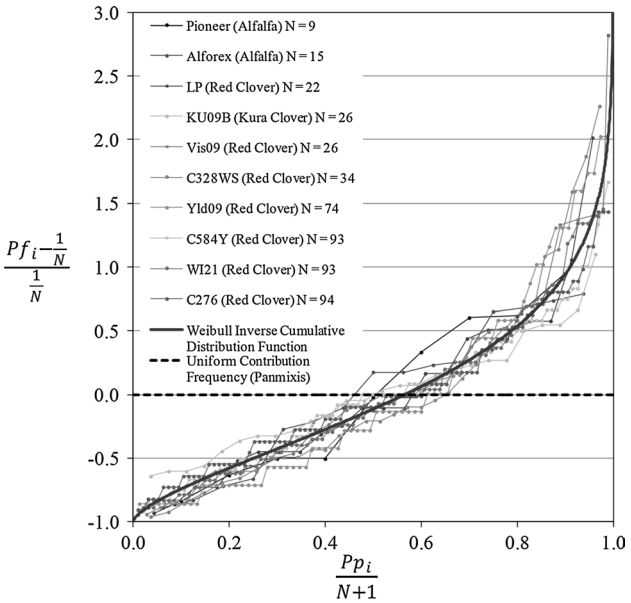

Fig. 2 Plots of polycross size standardized pollen-parent sequence position in array of all possible pollen-parents ordered from least frequent pollen-parent to most frequent pollen-parent (x axis) with panmixis standardized pollen-parent contribution to progeny frequencies (y axis) for 10 bee pollinated forage legume polycrosses ranging in size from 9 to 94 pollen-parents. Included is the fitted Weibull inverse cumulative distribution function ascertained using a mixed non-linear model approach ($y = -1 + 1.126(-\ln(1-x))^{\frac{1}{1.524}}, R^2 = 0.978$) and the expected panmixis standardized uniform distribution expected under panmixis or random mating

Fitting individual Weibull ICDF to each polycross revealed in most cases very good fits (R^2 0.918–0.991) with a narrower k (shape) parameter range (1.072–1.184) and a larger λ (scale) parameter range (1.118–2.196) (Table 4). A simple pooling of all the panmixis-scaled Pf_i values from all polycrosses yielded a very nice model (R^2 0.953, k 1.124, and λ 1.495) (Table 5). The problem with simple pooling was that each of the ten polycrosses were not weighted equally during model estimation; we therefore used a mixed model estimation approach with polycross as a random variable. A slightly better model was obtained (R^2 0.978, k 1.126, and λ 1.524) with individual polycross estimates, generally similar to models developed when each polycross was analyzed separately (Table 5).

The shape of the Weibull ICDF revealed that the least frequent contributing pollen-parents (bottom 25 % of pollen-parents) were contributing to progeny parentage at less than half the expected panmixis frequencies (Fig. 2). About 55 % of pollen-parents in the mid frequency contributing range were contributing from half of expected panmixis frequencies to one and one-half expected panmixis frequencies. Finally, the top 20 % of most fecund pollen-parents were in excess of one and one-half of expected panmixis frequencies with the top 10 and 5 % most fecund pollen-parents contributing at rates of two times to three times the expected panmixis frequencies respectively. In other words, based on our empirical observations, we consistently saw a few "super males" in the top percentiles of observed pollen-parent fecundity. These super fecund males based on our modeling are not an anomaly but are rather to be expected. The λ parameter is heavily influenced by the frequency of these super fecund males; this is one reason there was more variation in the λ parameter estimates between polycrosses. This was true particularly for the smaller polycrosses with <25 pollen-parents since in these cases few, if any, super fecund males are to be expected.

In order to utilize the model (5) we solved for Pf_i:

Table 5 Inverse Weibull cumulative distribution function parameter estimates in separate models for each polycross, a single model with all data pooled, and in a single mixed model with all data pooled but with polycross as a random variable, for panmixis scaled pollen-parent contribution frequency to progeny parentage ($\frac{Pf_i-\frac{1}{N}}{\frac{1}{N}}$) as the independent variable and with polycross size neutral scaled pollen-parent sequence position in array of possible pollen-parents ordered from Pf_i *min.* to Pf_i *max.* ($\frac{Pp_i}{N+1}$) as the dependent variable

Polycross name	No. of pollen-parents	Separate models					Combine mixed model				
		k	λ	SE k	SE λ	R^2	k[a]	λ[a]	SE k	SE λ	R^2
Pioneer	9	1.184	1.455	0.069	0.188	0.933	1.131	1.386	0.014	0.131	–
Alforex	15	1.172	1.788	0.051	0.189	0.918	1.133	1.718	0.016	0.111	–
LP08	22	1.125	1.312	0.036	0.070	0.964	1.127	1.288	0.010	0.104	–
KU09B	26	1.138	2.196	0.012	0.066	0.984	1.130	1.943	0.011	0.098	–
Vis09	26	1.144	1.397	0.026	0.057	0.976	1.130	1.357	0.011	0.103	–
C328WS	34	1.072	1.118	0.020	0.028	0.987	1.118	1.048	0.014	0.095	–
Yld09	74	1.097	1.227	0.018	0.027	0.977	1.114	1.151	0.016	0.087	–
C584Y	92	1.122	1.737	0.007	0.021	0.990	1.125	1.711	0.009	0.090	–
WI21	93	1.109	1.395	0.012	0.023	0.984	1.118	1.397	0.012	0.087	–
C276	94	1.132	1.759	0.007	0.021	0.991	1.130	1.719	0.009	0.092	–
Pooled $\frac{Pf_i-\frac{1}{N}}{\frac{1}{N}}$	Standardized	1.124	1.495	0.008	0.018	0.953	1.126	1.524	0.007	0.087	0.978

[a] Individual polycross estimates of k and λ and their standard errors were obtained from Proc Nlmixed random effects prediction to which parameter means were added

$$Pf_i = \frac{1}{N}\left[-1 + k\left(-\ln\left(1 - \frac{Pp_i}{N+1}\right)\right)^{\frac{1}{\lambda}}\right] + \frac{1}{N} \quad (6)$$

Specifically based on our empirical data, given $k = 1.126$ and $\lambda = 1.524$.

$$Pf_i = \frac{1}{N}\left[-1 + 1.126\left(-\ln\left(1 - \frac{Pp_i}{N+1}\right)\right)^{\frac{1}{1.524}}\right] + \frac{1}{N} \quad (7)$$

One exciting feature of the model is that to estimate Pf_i values requires only knowledge of polycross size N. One necessary aspect of model (7) is that the sum of Pf_i needs to equal one. Unfortunately this was not exactly the case for polycross sizes of 9–94 (the range of polycross sizes utilized in the study). The sum of Pf_i values for this polycross size range was 0.967–1.005 (Fig. 3). Going beyond this range the deviations from one were even greater for predicted Pf_i values, particularly for very small polycrosses (Fig. 3). We were unable to determine a model for this deviation from one in relationship to polycross size. Since our model was based on empirical observation, we felt it was acceptable to adjust model (7) so that the sum Pf_i had to equal one.

$$Pf_i^* = \frac{\frac{1}{N}\left[-1 + 1.126\left(-\ln\left(1 - \frac{Pp_i}{N+1}\right)\right)^{\frac{1}{1.524}}\right] + \frac{1}{N}}{\sum_{i=1}^{N} Pf_i} \quad (8)$$

where, Pf_i^* is the adjusted Pf_i such that $\sum_{i=1}^{N} Pf_i^* = 1$.

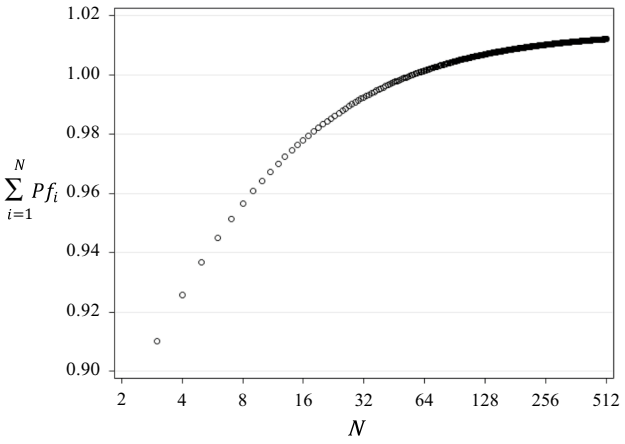

Fig. 3 Plot of polycross size (N) (x axis) with sum of predicted Pf_i frequencies for a polycross of size (N) based on model (7) (y axis). Individual frequencies within polycrosses do not exactly sum to one as necessary for frequency data

We expect that the adjusted Pf_i^* values are very close to the predicted Pf_i and that our adjustment maintains a Weibull distribution among our predicted values.

As part of a final analysis we examined predicted inbreeding that would occur using our model (F_W) versus inbreeding expected under panmixis (F_P). Assuming a synthetic plant breeding program syn 0 parents would constitute the initial mating (with a restricted set of parents) to produce syn 1 seed. The syn 1 seed would be increased to produce syn 2 seed in which generation the inbreeding would be evident (Busbice 1969). The syn 1 seed production field is usually large

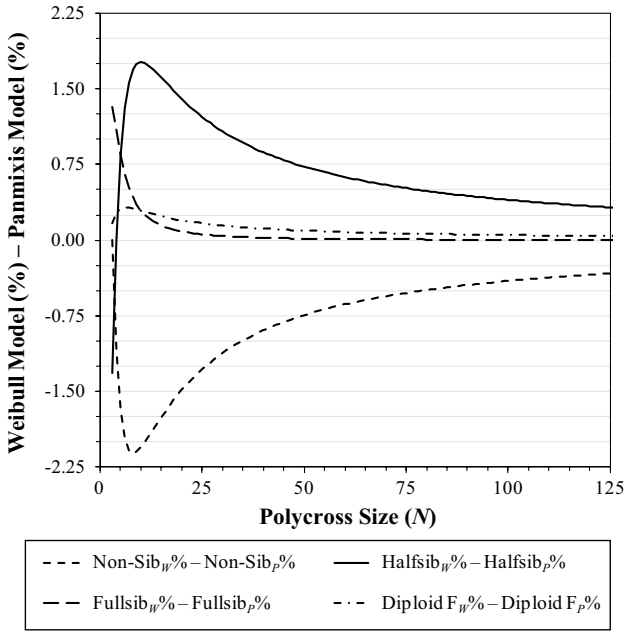

Fig. 4 Predicted Weibull distribution-based model (8) percent (sib_W %) minus predicted panmixis model percent (sib_P %) for expected non-sibs, halfsibs, and fullsibs relationship among progeny of a polycross size *N*. Assuming a diploid organism, the predicted percent inbreeding (*F*) for the Weibull distribution-based model (8) percent (F_W %) minus predicted panmixis model percent (F_P %) for different polycross sizes *N*

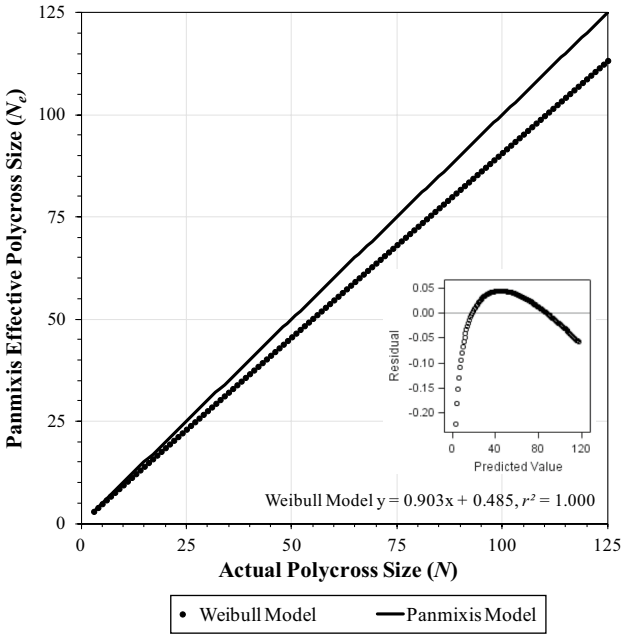

Fig. 5 Actual polycross size (*N*) (*x* axis) versus panmixis effective polycross size (N_e) for the Weibull distribution based model (8) and panmixis model. Insert graph plots predicted value with the residual of regression of $N_e = N\beta +$ intercept for the Weibull distribution based model (8)

with thousands of mating individuals. To estimate expected inbreeding in the syn 2 generation a coancestry approach was used (Malécot 1948; Busbice 1969) by first estimating expected syn 1 frequencies of fullsibs, halfsibs, and non-sibs using our model (8) ($fullsib_W$, $halfsib_W$, $non\text{-}sib_W$) and estimating these same frequencies under panmixis ($fullsib_P$, $halfsib_P$, $non\text{-}sib_P$) for a range of different syn 0 parent polycross sizes (Fig. 3). Syn 0 seed-parent contribution to the syn 1 generation was set as uniform. Assuming diploid genomes, syn 2 F_W % and F_P % were estimated. Inbreeding expressed as a percentage (F %) was used for convenience during graphing. Based on our empirical data model (8) showed greater inbreeding than panmixis (Fig. 4). Maximum deviation from panmixis was observed in polycross sizes at *N* equals 6 and 7 (Diploid F_W % − Diploid F_P % = +0.328 and +0.326 respectively). Following this maxima, the deviation decreases gradually. Initially both an excess of fullsibs and halfsibs compared to panmixis was observed in the syn 1 populations. This excess increased inbreeding levels in the syn 2 populations. In larger polycrosses ($N > 20$) it is primarily an excess of halfsibs that is apparent. The higher than expected syn 1 halfsib proportions led to higher than syn 2 expected panmixis inbreeding levels.

Finally, as plant breeders are familiar with discussing inbreeding effects in terms of polycross size, we estimated an effective polycross size (N_e) using the Weibull model (8) based on diploid F_P and F_W values (Fig. 5). For comparison we provide the panmixis model expectation which has a slope of one and an intercept of zero (Fig. 5). An almost perfect linear relationship was observed with a slope of 0.9. This slope indicates that effective N_e for bee-pollinated forage legume polycrosses is approximately 9 % smaller than the expectation under random mating.

Discussion

In this study we found that bee-pollinated forage legume polycrosses are clearly not random mating or in panmixis. Rather, all polycrosses showed the same predictable Weibull-distributed Pf_i values. This same pattern was observed among the two bee species used, which indicates that the distribution does not depend on the pollinator species. The same pattern was observed when DNA was extracted directly from seed, from greenhouse seedlings, or from greenhouse seedlings that had been transplanted to the field. This same pattern was observed if parental plants in the polycross were clonally replicated or not. Finally, the pattern was observed in polycrosses where seed-parents were uniformly and not uniformly sampled. All these observations led us to conclude that there is something inherent about the observed distribution. Expecting that the fecundity distribution is not uniform (i.e., violation of panmixis) is not a surprise based on the extensive body of

work in other plant species (Clegg 1980; Partridge 1983; Heywood 1986; Allard 1999; Nyquist and Santini 2007; Edelaar and Bolnick 2012; Jiang et al. 2013). What is surprising is, at least based on the inference space of this study, how easily the pollen-parent fecundity distribution is modeled knowing only the polycross size.

Right-skewed Pf_i value distributions observed in our study have been observed extensively by researchers measuring population life time fecundity distributions (Herrera and Jovani 2010; Chybicki and Burczyk 2013). A log-normal distribution is assumed for such fecundity hierarchy models (Solbrig and Solbrig 1984). Although in our study we observed a similar right-skewed distribution, among our populations the log-normal distribution was not the best fit. However, the premise that some form of fecundity hierarchy contributed to random mating may be the case. Fecundity hierarchies can occur due to genotypic plant size differences among mating individuals or due to environment-induced plant size differences resulting in a few larger plants with high fecundity and many smaller plants with low fecundity (Solbrig and Solbrig 1984). We did not collect data on plant size, however, because forage legumes populations are collections of heterozygous heterogeneous individuals; we have every reason to believe that some plant size differences or at least difference in number of flowers and by extension male fecundity would occur among mating individuals in the polycross [due to underlying genetic differences or due to random environmental chance (Heywood 1986)].

The inference space of this study is limited to physically small, isolated, bee-pollinated, outcross progeny forage legume polycrosses. It is unknown if our model has application to small, isolated, wind-pollinated species such as grasses or other mating situations such as among trees. It is unknown if pollinator density would alter the observed outcome. In our study, bee pollinator numbers were not a limiting factor to cross pollination since polycrosses were set up to maximize seed production. Further studies could examine if our model is applicable to forage legume polycross situations with varying pollinator availability or artificial pollination conducted by humans (e.g., chain crosses). Vleugels et al. (2014) report on a 111 parent polycross; their study allowed us to compare their observed results to our model prediction. In Vleugels et al. (2014) 1140 plants were planted in a 30 plant by 38 plant grid with 50 cm between plants (14.5 m × 18.5 m pollination space). After initial plant evaluation, 111 plants were retained in that study (plants were not moved from their position in the grid) and allowed to pollinate (pollinators were not identified). For the subset of top 10 seed-yielding plants, 139 progeny were identified as being crosses among these 10 top yielding plants (Vleugels et al. 2014 Table 3). Using these 139 progeny we calculated observed Pf_i values in Vleugels et al. (2014) and then estimated expected Pf_i values for a 10 parent

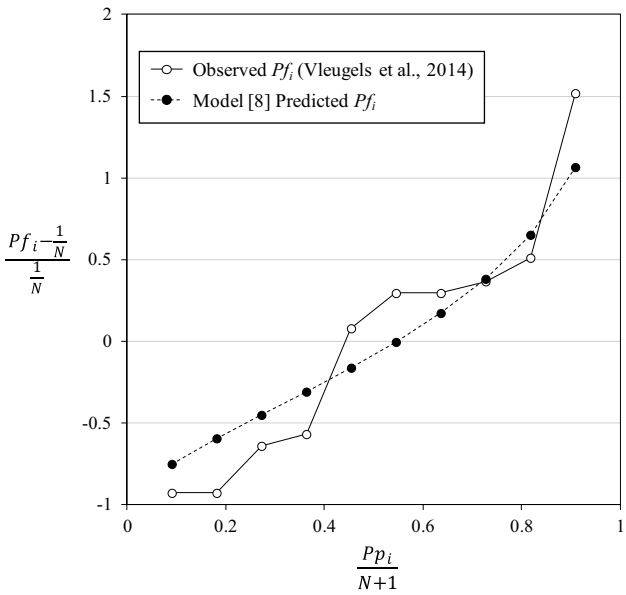

Fig. 6 Observed Pf_i values from Vleugels et al. (2014) compared to model (8) predicted Pf_i values ($R^2 = 0.881$, $P < 0.0001$)

polycross based on our model (8) (Fig. 6). A highly significant fit between the observed Pf_i values and our predicted values was observed ($R^2 = 0.881$, $P < 0.0001$). Notable in Vleugels et al. (2014) is that plants were spaced further apart, that spacing between pollinating-plants was irregular (due to culling), and that additional pollination was occurring simultaneously with the top 10 seed yielding plants which did not apparently alter the pollen-parent fecundity distribution among the top 10 seed-yielding plants. Despite these differences our model provided good predictions of the actual pollen-parent fecundity distribution.

A unique aspect of our study is that we transformed the pollen-parent fecundity data to be a deviation from panmixis and to be polycross-size neutral. Most previous work on fecundity distributions tries to take into account factors other than polycross size and therefore does not standardize pollen-parent fecundity data. However, Heywood (1986) provides N_e/N ratios when $F = 0$ based on data from 27 annual species. His data was based on seed set (i.e., seed-parent fecundity) and he found an average ratio of 0.42 with a range of 0.14 to 0.68. In our study, our model generates a N_e/N ratio of ~0.9 based on measured pollen-parent fecundity (Fig. 5). Heyward admits in his paper that due to sampling techniques his estimates may be underestimated. Still, there could be many reasons the two estimates are different such as: perenniality, differences in seed-parent and pollen-parent fecundity distributions, or because plants in the wild behave differently in terms of inter-plant competition and pollination than plants grown equally spaced with

controlled pollination in a plant breeding setting. Further studies could be pursued to compare seed-parent fecundity distributions to pollen-parent fecundity distributions to determine if they are the same or different within a species in a particular mating environment. Vleugels et al. (2014) published data suggests that in their case pollen-parent and seed-parent fecundity distributions may be different.

Even though we demonstrate a model for polycross mating, the impact on inbreeding, at least from the pollen-parent fecundity distribution perspective, is not particularly great (Fig. 4). Perhaps for many applications staying with panmixis assumption is adequate. However, when precise Pf_i values are desired or needed, using our model is more likely to replicate nature. As our model is empirical, improving its accuracy through additional observations would be advantageous, particularly in larger polycrosses to better define the λ parameter (e.g., polycross size 100–200). Sampling smaller polycrosses with less than nine individuals would be helpful since small-sized polycrosses were more difficult to incorporate into the models (Figs. 3, 5). Finally, since our model is empirical, ascertaining a theoretical basis for our observations would be desirable.

Author contribution statement H. Riday contributed field research, molecular genetic analysis, statistical analysis, and manuscript preparation. M.A. Smith contributed field research and input and critique during manuscript preparation. M.D. Peel contributed field research and input and critique during manuscript preparation.

Acknowledgments We would like to acknowledge the field technical support provided Rebecca Heidelberger and the molecular genetic laboratory support provided by John Raasch.

Conflict of interest This research was funded by USDA-ARS funds from CRIS projects 5090-21000-056-00 and 2080-21000-014-00. In-kind support was provided by the DuPont Pioneer alfalfa breeding program.

Appendix 1

```
PROC NLMIXED METHOD = FIRO;
  PARMS W = 1.5 K = 1.1 S2U1 = 0.01 CU1U2 = 0
  S2U2 = 0.01 S2E = 0.1;
  BOUNDS CU1U2 <=0, CU1U2 >=0;
  Y = −1 + ((K + U1)*(−1*LOG(1−RANGE))**
  (1/(W + U2)));
  MODEL FREQ ~ NORMAL(Y,S2E);
  RANDOM U1 U2 ~ NORMAL([0,0],[S2U1,CU1U2,S2
  U2]) SUBJECT = POLYCROSS;
RUN;
```

References

Allard RW (1999) Principles of plant breeding. Wiley, New York

Atugwu AI, Uguru MI (2012) Tracking fruit size increase in recombinants obtained from an interspecific cross between cultivated tomato (*Solanum esculentum*) and wild tomato relative (*S. pimpinellifolium*). J Plant Breed Crop Sci 4:62–71

Bartlerr AC, Butler GD (1975) Genetic control of the cabbage looper by a recessive lethal mutation. J Econ Entomol 68:331–335

Busbice TH (1969) Inbreeding in synthetic varieties. Crop Sci 9:601–604

Chybicki IJ, Burczyk J (2013) Seeing the forest through the trees: comprehensive inference on individual mating patterns in a mixed stand of *Quercus robur* and *Q. petraea*. Ann Bot 112:561–574

Clegg MT (1980) Measuring plant mating systems. Bioscience 30:814–818

Dodd ME, Silvertown J (2000) Size-specific fecundity and the influence of lifetime size variation upon effective population size in *Abies balsamea*. Heredity 85:604–609

Edelaar P, Bolnick DI (2012) Non-random gene flow: an underappreciated force in evolution and ecology. Trends Ecol Evol 27:659–665

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman Group Ltd, Harlow

Forbes C, Evans M, Hastings N, Peacock B (2011) Statistical distributions, 4th edn. Wiley, Hoboken

Globerson D, Genizi A, Staub JE (1987) Inheritance of seed weight in *Cucumis sativus* (L.) var. *sativus* and var. *hardwickii* (Royle) Kitamura. Theor Appl Genet 74:522–526

Gottlieb LD (1977) Genotypic similarity of large and small individuals in a natural population of the annual plant *Stephanomeria exigua* ssp. *Coronaria* (Compositae). J Ecol 65:127–134

Herrera CM, Jovani R (2010) Lognormal distribution of individual lifetime fecundity: insights from a 23-year study. Ecology 91:422–430

Heywood JS (1986) The effect of plant size variation on genetic drift in populations of annuals. Am Nat 127:851–861

Isobe S, Kölliker R, Hisano H, Sasamoto S, Wada T, Klimenko I, Okumura K, Tabata S (2009) Construction of a consensus linkage map for red clover (*Trifolium pratense* L.). BMC Plant Biol 9:57

Jiang Y, Bolnick DI, Kirkpatrick M (2013) Assortative mating in animals. Am Nat 181 **(E-Article)**

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kempthorne O (1969) An Introduction to genetic statistics. Iowa State University Press, Ames

Lambert RJ, Chung L-C (1995) Phenotypic recurrent selection for increased endosperm hardness in two high-lysine maize synthetics. Crop Sci 35:451–456

Malécot G (1948). Les mathématiques de l'hérédité. The mathematics of heredity, 1969 English translation by D.M. Yermanos, W.H. Freeman and Co., San Francisco

Nyquist WE, Santini JB (2007) Pollen dispersion within a population, nonrandom mating theory, and number of replications in polycross nurseries. Crop Sci 47:547–558

Partridge L (1983) Non-random mating and offspring fitness. In: Bateson P (ed) Mate choice. Cambridge University Press, Cambridge, pp 227–255

Rai KN, Talukdar BS, Singh SD, Rao AS, Rao AM, Andrews DJ (1994) Registration of ICMP 423 parental line of pearl millet. Crop Sci 34:1430

Riday H (2011) Paternity testing a non-linkage based marker assisted selection scheme for outbred forage species. Crop Sci 51:631–641

Riday H, Krohn AL (2010) Increasing population hybridity by restricting self-incompatibility alleles in red clover populations. Crop Sci 50:853–860

Riday H, Casler MD, Crooks A, Wood T (2007) Persistence of rotationally grazed red clover varieties grown in mixture with tall fescue in Wisconsin. Grass Clippings 2(1):3–6. Univ. Wisc. Ext.—Univ. Wisc.-Madison Center for Integrated Ag. Syst. and Col. of Ag. and Life Sci.—Univ. Wisc. Ag. Res. Sta

Riday H, Johnson DW, Heyduk K, Raasch JA, Darling MM, Sandman JE (2013) Paternity testing in an autotetraploid alfalfa breeding polycross. Euphytica 194:335–349

Riday H, Reisen P, Raasch JA, Santa-Martinez E, Brunet J (2015) Selfing rate in an alfalfa seed production field pollinated with leafcutter bees. Crop Sci. 55:1087–1095

SAS Institute Inc (2012) What's New in SAS® 9.3. SAS Institute Inc., Cary

Sato S, Isobe S, Asamizu E, Ohmido N, Kataoka R, Nakamura K, Kaneko T, Sakurai N, Okumura K, Klimenko I, Sasamoto S, Wada T, Watanabe A, Kohara M, Fujishiro T, Tabata S (2005) Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). DNA Res 12:301–364

Scheiner SM (1987) Size and fecundity hierarchies in an herbaceous perennial. Oecologia 74:128–132

Sledge MK, Ray IM, Jiang G (2005) An expressed sequence tag SSR map of tetraploid alfalfa (*Medicago sativa* L.). Theor Appl Genet 111:980–992

Solbrig OT, Solbrig DJ (1984) Size inequalities and fitness in plant populations. In: Dawkins R, Ridley M (eds) Oxford surveys in evolutionary biology, vol 1. Oxford University Press, Oxford, pp 141–159

Vleugels T, Cnops G, Roldán-Ruiz I (2014) Improving seed yield in red clover through marker assisted parentage analysis. Euphytica 200:305–320

Weibull W (1951) A statistical distribution function of wide applicability. J Appl Mech-Trans ASME 18:293–297

Weismann A (1883) On heredity. In: Essays upon heredity and kindred biological problems, 1889 English translation edited by E.B. Poulton, S. Schönland, and A.E. Shipley. Oxford University Press Warehouse, London, pp 69–105

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338