

# High-Resolution Single Nucleotide Polymorphism Genotyping Reveals a Significant Problem among Breeder Resources

Mitchell R. Lucas,\* Bao-Lam Huynh, Jeffery D. Ehlers, Philip A. Roberts, and Timothy J. Close

## Abstract

The logistics associated with a modern breeding program can be complex, relying on accuracy and communication between plant breeders, pathologists, quantitative geneticists, and support staff. International and academic facets may bring additional challenges to already error prone activities including the development, maintenance, and distribution of lines. Furthermore, practices such as bulking of seed and the maintenance of within-accession variation among landraces must be considered when pursuing marker-assisted approaches to breeding.

**C**ULTIVARS, GERMPLASM, AND POPULATIONS that have been bred by specific design have several expected characteristics including allelic diversity, heterozygosity, and individuality. The existence of rogues, individuals that violate these premises, is documented among important crop and model species. Authors often do not elaborate on the potential origin of rogues, but in some cases hypotheses have been formed with outlandish biological explanations. From a practical standpoint the unintentional use of rogues can be problematic when used for breeding or when developing breeder resources. Undetected rogues may also have financially costly impacts if they are included in field trials or are repeatedly genotyped.

Fortunately, insight provided by high-throughput genotyping can assess how well an individual matches its pedigree record. The original intent of genotyping resources was not for error detection “forensics”; however, in actual practice the benefits are immediate and significant. Here we discuss typical examples of rogues, their impacts, and detection using data from our work on cowpea [*Vigna unguiculata* (L.) Walp.].

## Origin and Detection of Rogues

Unintentional outcrossing events are likely sources of rogues. Even species that normally have high frequencies of self-pollination will occasionally outcross (Lloyd and Schoen, 1992). These outcrosses can lead to heterozygosity and the presence of nonparental alleles. Organizational errors (i.e., harvesting seed from volunteer plants, mislabeling, different paths of single-seed descent, and/

Published in The Plant Genome 6.  
doi: 10.3835/plantgenome2012.08.0020  
© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

M.R. Lucas, J.D. Ehlers, and T.J. Close, Dep. of Botany and Plant Sciences, Univ. of California-Riverside, Riverside, CA 92521; B.-L. Huynh and P.A. Roberts, Dep. of Nematology; Univ. of California-Riverside, Riverside, CA 92521. Received 10 Aug. 2012.  
\*Corresponding author (Mitchell.lucas@email.ucr.edu).

**Abbreviations:** QTL, quantitative trait loci/locus; SNP, single nucleotide polymorphism.

**Table 1. Excerpt from Lucas et al. (2011) describing rogues among 13 mapping populations of cowpea determined from the analysis of 1536-plex single nucleotide polymorphism genotype information.**

Population	Individuals genotyped	Individuals used for mapping	HNPG <sup>†</sup>	Genotypically identical sets of individuals
CB27 × IT97K-556-6	95	92	1	2
CB27 × IT82E-18	166	160	2	4
CB27 × UCR 779	58	56	0	2
CB46 × IT93K-503-1	130	114	16	0
524B × IT84S-2049	91	85	5	1
Dan Ila × TVu-7778	113	79	11	23
Yacine × 58-77	141	97	43	1
Sanzi × Vita 7	142	122	11	9
IT84S-2246 × IT93K-503	93	88	5	0
IT84S-2246 × Mouride	92	87	5	0
TVu14676 × IT84S-2246-4	147	136	10	1
CB27 × 24-125B-1	108	87	18	3
LB30#1 × LB1162 #7	95	90	4	1

<sup>†</sup>The number of individuals that are highly heterozygous or nonparental in genotype (HNPG).

or mixing seed from different lines) could also create rogues, including duplicate lines in a collection.

All 13 populations used to construct the consensus genetic map of cowpea (Lucas et al., 2011) contained at least some rogues (Table 1), which range from 3 to 31% providing an average of 11.3% rogue. In our experience, these errors have been more common than expected. But a search of the literature provides a fair number of documented cases of such problems. Genotype information has been used to identify duplicated lines of apple (*Malus domestica* Borkh.) (Hokanson et al., 1998), orange [*Poncirus trifoliata* (L.) Raf.] (Fang et al., 1997), and rice (*Oryza sativa* L.) (Virk et al., 1995). In our work, 10 out of the 11 recombinant inbred populations genotyped for 1536 single nucleotide polymorphisms (SNPs) were found to contain duplicate individuals (Table 1). In that work and in subsequent quantitative trait loci (QTL) analyses (Lucas et al., 2012a, 2012b) duplicate lines and other rogues were omitted from data analysis. Interestingly, cowpea lines that are identical often have a sequential or similar name (Supplemental Table S1). Seventeen out of 54 instances of duplications occurred between lines of sequential naming (i.e., line -036 identical to line -037) while 14 duplications were found between lines with a similar name (i.e., line -049 identical to line -094, line -088-2 identical to line -002, and line -084 identical to line -184). Such duplications seem most likely to be the result of human error that may have occurred at a number of different stages during the creation, maintenance, or distribution of these inbred populations.

Within-accession variation has been diagnosed via molecular markers in rice (Olufowote et al., 1997), spanish melon (*Cucumis melo* L.) (Lopez-Sese et al., 2002), and cowpea (Hearne et al., 2010). A genome covering set of 80 simple sequence repeats was recently used to verify pedigree records in apple (*Malus* spp.) (Evans et al., 2011).

**Table 2. Number of single nucleotide polymorphism (SNP) polymorphisms found by genotyping two lines with the same name from six accessions of cowpea. The distribution of polymorphisms is indicated among the 11 linkage groups of the cowpea genome.**

Accession	Number of polymorphic loci <sup>†</sup>	Location of polymorphisms on linkage groups		
		Beginning	End	Dispersed
Yacine	10	6	7	–
TVu-16722	63	2 and 5	1	3 and 9
TVu-15112	90	1, 2, 3, 10, and 11	–	5, 7, and 9
TVu-10513	103	1, 6, and 9	3 and 8	2, 4, 5, and 10
58-77	113	3, 4, 5, and 8	–	1, 2, 10, and 11
TVu-14321	159	3 and 8	4	1, 2, 5, 9, and 10

<sup>†</sup>With respect to a 1536-plex SNP genotyping platform.

In addition to the identification of 15 rogues, two plants known by the same name (Priscilla) were found to be different based on genotype information collected in that study. We assessed the possibility of cowpea within-accession variation by genotyping inbred stocks carrying the same name but provided from different sources of seed. This was performed to capture one aspect of within-accession variation, different paths of single-seed descent rather than assessing heterogeneity of one seed stock. Although some accessions were identical between seed sources, based on 1536-plex SNP genotyping, six sets of inbreds were identified that were different at many loci (Table 2). These polymorphisms between lines with the same name were only dispersed among some linkage groups and tend to be localized to the ends of linkage groups. These haplotype blocks provide evidence for divergent descent from a common parent.

Molecular markers can also be used to assess the frequency of self-fertilization. Single nucleotide polymorphisms are being considered for a molecular hybridity test in faba bean (*Vicia faba* L.) (Cottage et al., 2012). In that work 31 out of 32 plants self-fertilized in an overwinter glass house while one outcrossed. A similar situation was observed among the SNP genotyped populations of cowpea. Six out of the 11 recombinant inbred populations of cowpea used to build the consensus map contained lines that were heterozygous far beyond expectation based on the number of inbreeding generations (Supplemental Table S2), often correlating with the presence of nonparental alleles. This type of variation may arise from unintentional outcrossing or organizational errors or when inherently variable land race accessions are used as parents.

As indicated above, genotype information can also identify carriers of nonparental alleles, similar to what was observed in wheat (*Triticum aestivum* L. and *Triticum turgidum* L.) (Khan et al., 2000), banana (*Musa* spp.) (Crouch et al., 1999), and rapeseed (*Brassica napus* L.) (Trick et al., 2009). During the construction of the cowpea consensus map, nine out of the 11 recombinant inbred populations were found to have at least one individual carrying nonparental alleles (Lucas et al.,

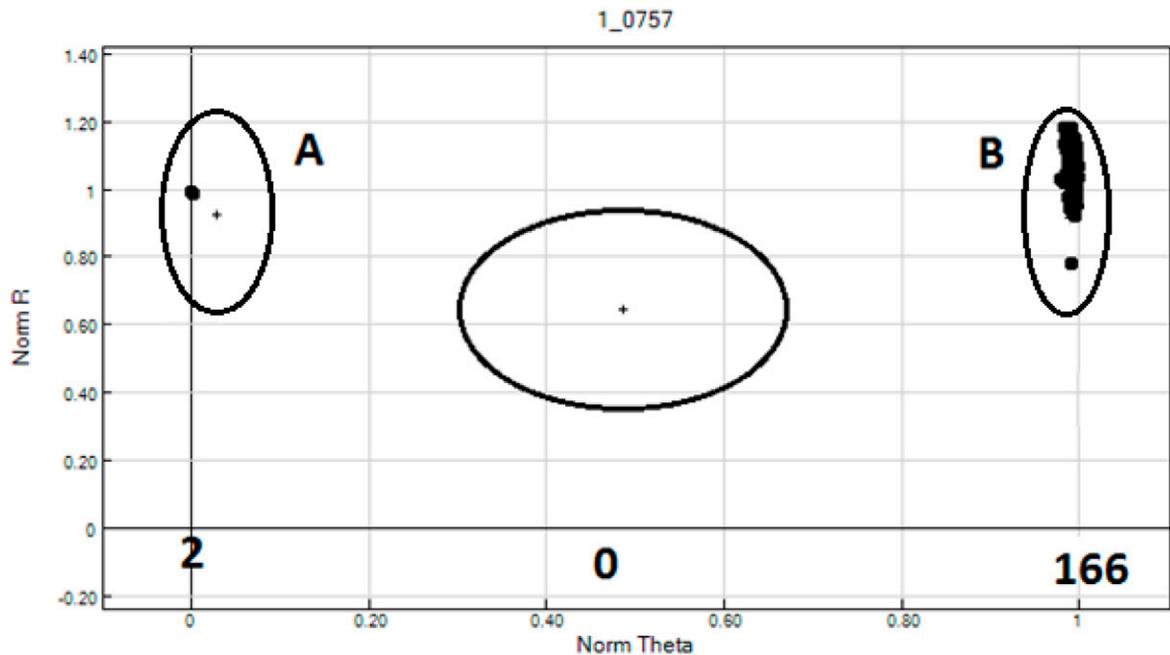


Figure 1. Two individuals with nonparental genotype calls among a recombinant inbred population of cowpea. Individuals homozygous for allele A (A) and allele B (B) at single nucleotide polymorphism locus 1\_0757 are shown using Illumina GenomeStudio genotype visualization software (Illumina, 2010). The parents and 164 progeny are monomorphic and contain allele B while two individuals are homozygous for allele A.

2011). In that work a fixed array genotyping platform was used that provided information for some markers that were fixed among the parents. Rogues were quickly identified when working with genotype information graphically (Fig. 1). As a part of this communication we detail nonparental allele containing rogues among the populations used to construct the cowpea consensus map (Supplemental Table S2).

A notorious example of apparent outcrossing in *Arabidopsis thaliana* (L.) Heynh. has led to controversial discussion. Lolle et al. (2005) provide a biological explanation for the genomewide inheritance of nonparental alleles that involves a hypothetical cache of ancestral ribonucleic acid, a notion that has no support from any prior work in any organism. Alternative explanations crafted on the basis of little experimental data are also being considered (Ray, 2005; Chaudhary, 2005; Comai, 2005; among many others); however, we underline the elegant work of Peng et al. (2006) that provides the most parsimonious explanation concerning the origin of nonparental alleles, unintentional outcrossing.

### Impact of Rogues

Rogues are particularly problematic when taking marker-assisted approaches to breeding. Accurate genetic maps and statistical estimates are required to associate markers with traits. Maps are often constructed by observing recombination frequencies among members of a population. Progeny provide a sample of possible recombination events occurring during meiosis and no two individuals should be identical. Additionally, rogues carrying nonparental alleles may contribute phenotypic

variation disregarded by statistical models intended to operate on populations of biparental design. Lines carrying nonparental alleles are known to be significant obstacles for map construction (Ming et al., 1997; Lucas et al., 2011). Inaccurate estimations of genetic distances and associations may confound attempts aiming to use a marker-assisted approach to selection. To assess the impact of rogues on developing resources for breeding, we built genetic maps and performed QTL analysis with and without rogue individuals. Maps constructed with and without rogues are different (Fig. 2) and often have greater distances between bins when rogues are included. Quantitative trait loci analyses are also affected (Fig. 3), where the inclusion of rogues can suggest the existence of a QTL where none are known to exist. Furthermore, rogues with excessive heterozygosity or nonparental alleles that perform strong phenotypically may be incorrectly known as inbred while these individuals are actually benefiting from hybrid vigor or alleles thought to be absent from its pedigree. The unintentional use of these individuals for breeding or for the development of marker-trait associations may lead to unpredictable outcomes including multiform progeny and linkage drag.

Population size is a major constraint for marker-assisted breeding initiatives because of financial costs associated with genotyping and phenotyping. Selection decisions and marker-trait associations are typically developed by observing the performance of lines in replicated, multilocation field trials. The logistics associated with these operations demand a substantial proportion of financial resources available to a breeding program. Removing rogues before phenotyping trials

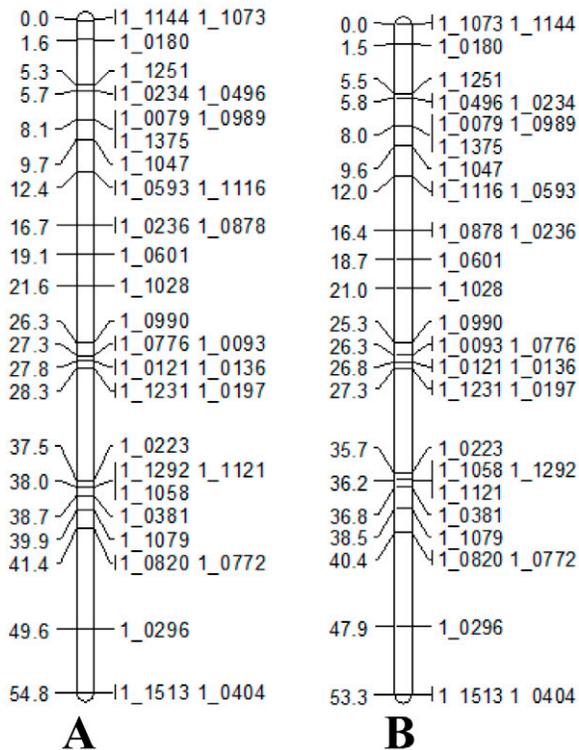


Figure 2. Linkage maps constructed with (A) and without (B) rogues.

and extensive genotyping would eliminate unnecessary expenses. Therefore, rogues are not only problematic but also financially inefficient to maintain.

Geneticists should put a greater emphasis on developing high quality populations rather than relying on historical populations made by breeders who may be more permissive of rogues. Improved analyses, community resources, and financial efficiency could be realized by approaching genotype data from a forensics perspective. This approach could also be used as a quality control measure for the future development of lines.

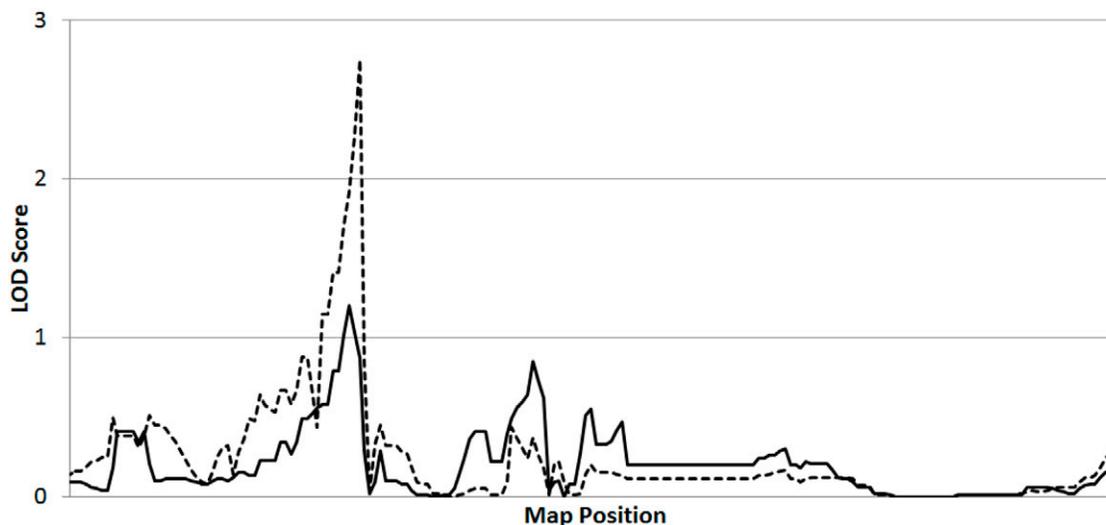


Figure 3. Logarithm of the odds (LOD) score traces for quantitative trait loci analyses performed with (solid line) and without (dashed line) rogues.

Our group is verifying pedigree records for members of a multiparent advanced generation intercross (MAGIC) population by genotyping the progeny for genomewide markers known to segregate among the intended parents. This work has helped verify cross-pollination events and helps to ensure a quality community resource is being provided on the basis of genotype-validated seed. Validation via genotyping would be particularly valuable for species that are difficult to cross or for lines or populations that will be heavily used.

### Supplemental Information Available

Supplemental material is included with this manuscript.

### References

- Chaudhary, A. 2005. Plant genetics: Hothead healer and extragenomic information. *Nature* 437:E1. doi:10.1038/nature04062
- Comai, L. 2005. A toxic mutator and selection alternative to the non-Mendelian RNA cache hypothesis for hothead reversion. *Plant Cell* 17:2856–2858. doi:10.1105/tpc.105.036293
- Cottage, A., K. Gostkiewicz, J.E. Thomas, R. Borrows, A.M. Torres, and D.M. O’Sullivan. 2012. Heterozygosity and diversity analysis using mapped single nucleotide polymorphism in a faba bean inbreeding programme. *Mol. Breed.* 30:1799–1809. doi:10.1007/s11032-012-9745-4
- Crouch, J.H., H.K. Crouch, H. Constandt, A. Van Gysel, P. Breynne, M. Van Montagu, R.L. Jarret, and R. Ortiz. 1999. Comparison of PCR-based molecular marker analyses of *Musa* breeding populations. *Mol. Breed.* 5:233–244. doi:10.1023/A:1009649521009
- Evans, K.M., A. Patocchi, F. Rezzonico, F. Mathis, C.E. Durel, F. Fernandez-Fernandez, A. Boudichevskaia, F. Dunemann, M. Stankiewicz-Kosyl, L. Gianfranceschi, M. Komjanc, M. Lateur, M. Madduri, Y. Noordijk, and W.E. van de Weg. 2011. Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: Trueness-to-type of cultivars and their parentages. *Mol. Breed.* 28:535–547. doi:10.1007/s11032-010-9502-5
- Fang, D.Q., M.L. Roose, R.R. Krueger, and C.T. Federici. 1997. Fingerprinting trifoliolate orange germ plasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theor. Appl. Genet.* 95:211–219. doi:10.1007/s001220050550
- Hearne, S., J. Franco-Duran, and E. Magembe. 2010. Optimal units of selection – How many plants are representative of diversity in collections of cowpea landrace accessions? In: *Proceedings of the*

- World Cowpea Conference, Saly, Senegal. 27 Sept.–1 Oct. 2010. IITA, Ibadan, Nigeria. p. 46–47.
- Hokanson, S.C., A.K. Szewc-McFadden, W.F. Lamboy, and J.R. McPerson. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* × *domestica* Borkh. core subset collection. *Theor. Appl. Genet.* 97:671–683. doi:10.1007/s001220050943
- Illumina. 2010. GenomeStudio genotyping module v.2010.3. Illumina, Inc., San Diego, CA. [http://www.illumina.com/software/genomestudio\\_software.ilmn](http://www.illumina.com/software/genomestudio_software.ilmn).
- Khan, I.A., J.D. Procnier, D.G. Humphreys, G. Tranquilli, A.R. Schlatter, S. Marcucci-Poltri, R. Frohberg, and J. Dubcovsky. 2000. Development of PCR-based markers for a high grain protein content gene from *Triticum turgidum* ssp. *dicoccoides* transferred to bread wheat. *Crop Sci.* 40:518–524. doi:10.2135/cropsci2000.402518x
- Lloyd, D.G., and D.J. Schoen. 1992. Self- and cross-fertilization in plants. I. Functional dimensions. *Int. J. Plant Sci.* 153:358–369. doi:10.1086/297040
- Lolle, S.J., J.L. Victor, J.M. Young, and R.E. Pruitt. 2005. Genome-wide non-Mendelian inheritance of extra-genomic information in *Arabidopsis*. *Nature* 434:505–509. doi:10.1038/nature03380
- Lopez-Sese, A.I., J. Staub, N. Katzir, and M.L. Gomez-Guillamon. 2002. Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD and SSR markers to assess strategies for large collection evaluation. *Euphytica* 127:41–51. doi:10.1023/A:1019904224170
- Lucas, M.R., N.N. Diop, S. Wanamaker, J.D. Ehlers, P.A. Roberts, and T.J. Close. 2011. Cowpea-soybean synteny clarified through an improved genetic map. *Plant Gen.* 4:218–225. doi:10.3835/plantgenome2011.06.0019
- Lucas, M.R., J.D. Ehlers, B.L. Huynh, N.N. Diop, P.A. Roberts, and T.J. Close. 2012a. Markers for breeding heat tolerant cowpea. *Mol. Breed.* doi:10.1007/s11032-012-9810-z
- Lucas, M.R., J.D. Ehlers, P.A. Roberts, and T.J. Close. 2012b. Markers for quantitative inheritance of resistance to foliar thrips in cowpea. *Crop Sci.* 52:2075–2081. doi:10.2135/cropsci2011.12.0684
- Ming, R., J.L. Brewbaker, R.C. Pratt, T.A. Musket, and M.D. McMullen. 1997. Molecular mapping of a major gene conferring resistance to maize mosaic virus. *Theor. Appl. Genet.* 95:271–275. doi:10.1007/s001220050559
- Olufowote, J.O., Y. Xu, X. Chen, W.D. Park, H.M. Beachell, R.H. Dilday, M. Goto, and S.R. McCouch. 1997. Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome* 40:370–378. doi:10.1139/g97-050
- Peng, P., S.W.L. Chan, G.A. Shah, and S.E. Jacobsen. 2006. Plant genetics: Increased outcrossing in hothead mutants. *Nature* 443:E8. doi:10.1038/nature05251
- Ray, A. 2005. Plant genetics: RNA cache or genome trash? *Nature* 437:E1–E2. doi:10.1038/nature04063
- Trick, M., Y. Long, J. Meng, and I. Bancroft. 2009. Single nucleotide polymorphism (SNP) discovery in the polyploidy *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol. J.* 7:334–346. doi:10.1111/j.1467-7652.2008.00396.x
- Virk, P.S., H.J. Newbury, M.T. Jackson, and B.V. Ford-Lloyd. 1995. The identification of duplicate accessions within a rice germplasm collection using RAPD analysis. *Theor. Appl. Genet.* 90:1049–1055. doi:10.1007/BF00222920