Roundtable

Detecting (trans)gene flow to landraces in centers of crop origin: lessons from the case of maize in Mexico

David A. CLEVELAND[1]*,**, Daniela SOLERI[1]**, Flavio ARAGÓN CUEVAS[2], José CROSSA[3] and Paul GEPTS[4]

[1] Environmental Studies Program, University of California, Santa Barbara, CA 93106-4160, USA

[2] Instituto Nacional de Investigaciones Forestales y Agropecuarias (INIFAP), Campo Experimental Valles Centrales,
Melchor Ocampo No. 7, Santo Domingo Barrio Bajo, Etla, Oaxaca, México

[3] Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641, México D.F., México

[4] Department of Plant Sciences/MS1, Section of Crop and Ecosystem Sciences, University of California, Davis, CA 95616-8780, USA

There is much discussion of the probability of transgene flow from transgenic crop varieties to landraces and wild relatives in centers of origin or diversity, and its genetic, ecological, and social consequences. Without costly research on the variables determining gene flow, research on transgene frequencies in landrace (or wild relative) populations can be valuable for understanding transgene flow and its effects. Minimal research requirements include (1) understanding how farmer practices and seed systems affect landrace populations, (2) sampling to optimize N_e/n (effective /census population size), (3) minimizing variance at all levels sampled, and (4) using N_e to calculate binomial probabilities for transgene frequencies. A key case is maize in Mexico. Two peer-reviewed papers, based on landrace samples from the Sierra Juárez region of Oaxaca, Mexico, reached seemingly conflicting conclusions: transgenes are present (Quist and Chapela, 2001, *Nature* **414**: 541–543; 2002, *Nature* **416**: 602) or "detectable transgenes" are absent (Ortiz-García et al., 2005, *Proc. Natl. Acad. Sci. USA* **102**: 12338–12343 and 18242). We analyzed these papers using information on Oaxacan maize seed systems and estimates of N_e. We conclude that if Quist and Chapela's results showing presence are accepted, Ortiz-García et al.'s conclusions of no evidence of transgenes at detectable levels or for their introgression into maize landraces in the Sierra de Juárez of Oaxaca are not scientifically justified. This is because their samples are not representative, and their statistical analysis is inconclusive due to using n instead of N_e. Using estimates of N_e based on Ortiz-García et al.'s n, we estimate that transgenes could be present in maize landraces in the Sierra Juárez region at frequencies of ~1–4%, and are more likely to be present in the 90% of Oaxacan landrace area that is not mountainous. Thus, we have no scientific evidence of maize transgene presence or absence in recent years in Mexico, Oaxaca State, or the Sierra Juárez region.

Keywords: effective population size / sampling / gene flow / Mexico / maize / transgenes / genetically engineered crops / landraces / traditional agriculture / third world / Oaxaca / research design

Las consecuencias genéticas, ecológicas y sociales sobre el posible flujo de genes de cultivos transgénicos hacia especies autóctonas y sus relativos silvestres han estado recientemente en discusión. Sin investigaciones costosas sobre los variables que afectan la frecuencia del flujo genético, investigaciones sobre la frecuencia de transgenes en poblaciones de cultivos autóctonas y sus parientes silvestres pueden ser muy útil. Los requerimientos mínimos de investigación incluyen (1) entender las practicas de los agricultores y como los sistemas de semilla afectan las poblaciones de especies autóctonas, (2) muestreo para optimizar N_e/n (tamaño efectivo/tamaño censo), (3) minimizar la varianza a todos los niveles de muestreo, (4) usar el tamaño efectivo, N_e, para calcular la probabilidad de la frecuencia de transgénicos (usando la distribución binomial). Un ejemplo clave es lo de maíz en México. Dos artículos basados en el muestreo de razas de maíz local (criollo) en la región de la Sierra Juárez de Oaxaca, México dan conclusiones aparentemente contradictorias: hay transgénicos

---

* Corresponding author: cleveland@es.ucsb.edu
** These authors contributed equally to this work.

(Quist y Chapela, 2001, *Nature* **414**: 541–543; 2002, *Nature* **416**: 602) o no hay "transgénicos detectables" (Ortiz-García et al., 2005, *Proc. Natl. Acad. Sci. USA* **102**: 12338–12343 y 18242). Nosotros hemos analizado estos artículos usando información de los sistemas de semilla de Oaxaca y cálculos de N_e. Concluimos que si los resultados de Quist y Chapela mostrando la presencia de transgénicos son aceptados, las conclusiones de Ortiz-García et al. de que no existen transgénicos detectables en los maíces de la Sierra Juárez de Oaxaca no se justifican científicamente. Esto se debe a que los tamaños de muestra usados por Ortiz-García no son representativos y su análisis estadístico no es concluyente por que usaron n en lugar de N_e. Usando estimadores de N_e basado en el n de Ortiz-García et al., nosotros estimamos que transgénicos pueden ser presente en las razas de maíz de la región de Sierra Juárez en frecuencias de ~1–4%, además es posible que estén presentes en el 90% del área sembrado con maíz criollo que se encuentra en las zonas no montañosas de Oaxaca. Por lo tanto no hay evidencias científicas de la presencia o ausencia de maíz transgénico en años recientes en México, o en el estado de Oaxaca, o en la región de la Sierra Juárez.

INTRODUCTION

There has been much scientific and popular discussion of the potential for transgene flow – whether by seed or pollen – from transgenic (genetically engineered) crop varieties (TGVs) to non-transgenic varieties and wild and weedy relatives of crops, and its potential effects on genetic diversity, ecology and society (CEC, 2004; NRC, 2002; NRC, 2004). Gene flow is a multi-step process, beginning with seed flow and pollen flow, followed by hybridization, and introgression (the incorporation and stable inheritance of introduced alleles, or genes in the case of transgenes). There is now ample evidence that hybridization among modern varieties, landraces, and wild or weedy relatives takes place for many crop species (Ellstrand, 2003). In the case of transgenes, for example, they have been backcrossed from transgenic sunflower into a wild relative (both *Helianthus annus*) (Snow et al., 2003), and have hybridized from transgenic oilseed rape (*Brassica napus*) into nontransgenic and transgenic oilseed rape (Hall et al., 2000). The results of hybridization and introgression are difficult to predict, but they could be negative, including a reduction of crop genetic diversity (Berthaud and Gepts, 2004; Ellstrand, 2003; Gepts and Papa, 2003; Papa et al., 2006; Soleri et al., 2006).

Whether the process of gene flow begins at all, how far it goes, and what effects it has, depends on each specific combination of genotypes (TGVs, non-transgenic varieties, landraces, wild relatives) and environments involved, including farmer management practices, crop reproductive biology, spatial distribution of genotypes, phenological synchrony, and absolute and relative fitness of TGVs and TGV × landrace hybrids and subsequent generations (Cleveland and Soleri, 2005; Ellstrand, 2003; Gepts and Papa, 2003). For example, it has been assumed, based on interpretations of population genetics theory, that transgenes would not introgress into landraces and persist if the selection pressure under which they were designed to function was absent. However, this may not always be the case – backcrosses to *Brassica rapa* from glufosinate-resistant transgenic *B. napus* × *B. rapa* hybrids (Snow et al., 1999) in one case, and, as mentioned earlier, a wild sunflower with a *Bt* gene backcrossed into it in another (Snow et al., 2003) showed no fitness costs associated with transgene presence when no selection for the transgene was exerted.

In the absence of more costly research on the variables determining seed and pollen flow, hybridization and introgression in each specific situation, research on transgene frequencies in landrace (or wild relative) populations over space and time can be valuable for understanding the determinants of transgene flow, and may also provide important clues about its effects. The need for data increases as pressure for approval of TGVs of food crops for field production in their centers of origin and diversity increases amid concerns about potential effects of transgene flow, as with rice in China (Huang et al., 2005; Lu and Snow, 2005; Zi, 2005).

However, accurate and reliable estimates of transgene presence need to be based on an understanding of the structure and dynamics of local landrace populations and how farmers manage them, and on sampling and estimates of effective population size appropriate for these populations. We describe the general principles involved, illustrating these for the case of maize in Mexico, and analyze the two peer-reviewed papers on transgene presence in Mexican maize landraces – papers with seemingly opposite results.

THE CASE OF MAIZE IN MEXICO

Mexico is the center of origin and a center of diversity of maize (Matsuoka et al., 2002). There are many reasons to hypothesize transgene presence in Mexican maize landraces, including a highly outcrossing reproductive

biology, diverse agronomic practices, extensive maize plantings, seed procurement and selection practices, farmer experimentation, and large US maize grain imports (Cleveland and Soleri, 2005). In addition to a wealth of diverse traditional landraces grown by Mexican smallholder farmers, populations of teosinte, the ancestor of maize (*Zea mays* subsp. *parviglumis*) grow there as well. Extensive gene flow, including introgression, occurs in Mexico between maize landrace populations (Pressoir and Berthaud, 2004), and non-transgenic modern varieties and landraces (Serratos et al., 1997), though the effects of this have not been studied. Hybridization and introgression occur in Mexico at low rates between maize and teosinte (Baltazar et al., 2005). There is also recent experimental evidence based on artificial crosses that introgression can occur between TGVs and Mexican landraces (Ellstrand, personal communication, November 2005). Maize is also very important socially and culturally in that country (Esteva and Marielle, 2003). In July 1998 Mexico placed a moratorium on experimental and commercial planting, but not the consumption of transgenic maize (Alvarez-Morales, 2000). Hence, large quantities of maize grain from the US continued to be imported, including white maize for human consumption, which until 2003 was distributed by the Mexican government as whole grain directly to consumers. These shipments very likely include many transgenic seeds (Cleveland and Soleri, 2005), and many farmers have planted food grain as seed (Soleri et al., 2005).

Thus, there are many reasons to hypothesize that transgenes are present in Mexican maize landraces. However, transgene presence in any specific situation cannot be assumed and depends on the factors described above, about which little is known for the many diverse environments of Mexico (Baltazar and Schoper, 2002; Cleveland and Soleri, 2005; Gepts and Papa, 2003; Ma et al., 2004). Highly controversial Mexican national legislation passed in February 2005 prepares the way for the possible approval of experimental and commercial planting of transgenic maize (Ballinas and Becerril, 2005; Peregrina and Crúz, 2005). Therefore, the need for data is great, and the two peer-reviewed papers on transgenes in Mexican maize landraces are valuable contributions to our knowledge.

The first paper by Quist and Chapela (2001), and a subsequent response to critiques (Quist and Chapela, 2002), reported evidence of the presence of three transgene motifs from *Bt* maize (*Bt*, CaMV promoter and NOS terminator) in maize landraces collected in fall 2000 in the Municipality of Ixtlán de Juárez in the state of Oaxaca in southern Mexico. They pooled kernels of each of the 6 ears collected separately, reported 4 ears contained transgenic DNA (Quist and Chapela, 2001), and estimated ~1% of kernels from each of these were transgenic (based on the CaMV promoter) (Quist and Chapela, 2002). (Both Quist and Chapela and Ortiz-García et al. used the term "cob" which is the term for the rachis and glumes without seeds – here we use "ear", the term for the cob with seeds.) It was the first peer-reviewed report of transgene presence in landraces of crops in a center of origin and diversity and, therefore, greatly increased the intensity of the debate about GE crops. Quist and Chapela's conclusion that transgenes had introgressed into maize landraces has been criticized (Kaplinsky et al., 2002; Metz and Fütterer, 2002), but their conclusion of transgene presence in landraces via hybridization was corroborated by subsequent Mexican government studies (Alvarez-Morales, 2002), although those were not published in peer-reviewed journals. By spring 2002, most observers, including the majority of critics of the Quist and Chapela paper, appeared to accept transgene flow into maize landraces as inevitable – "DNA flies around all over the place down on the farm" and is "as normal and natural, well, as agriculture itself" (Nat. Biotechnol., 2002).

Almost four years later, Ortiz-García et al. (2005a) reported absence of evidence of transgenes in maize landraces collected in 2003 and 2004 in the same area of Oaxaca as Quist and Chapela's study (with minor corrections reported in Ortiz-García et al., 2005b). Accepting the null hypothesis of absence of transgenes at detectable frequencies, as Ortiz-García et al. did, is potentially more open to question and misinterpretation (Andow, 2003) than accepting the alternative hypothesis of presence of transgenes, as Quist and Chapela did. Not only is sampling more challenging, there is also the danger of confounding absence of evidence with evidence of absence (Altman and Bland, 1995). We focus here on the Ortiz-García et al. study, specifically on seeds they collected in 2004, since they reported the lowest probability of failing to detect transgenes for this sample. We analyze the representativeness of their plant materials based on sampling methods, and their statistical analysis, using population genetics principles and information on the dynamics of local maize populations and farmer management. Throughout our analysis, we use the most conservative estimates, *e.g.*, those most supportive of Ortiz-García et al. Our goal is to explore scientifically sound approaches to estimating (trans)gene frequencies in landraces, and contribute to an open and constructive discussion of the general issues, and of the case of maize transgenes in Mexico, in particular.

SAMPLING

If the objective of sampling for transgenes in landrace populations is to determine *presence* or *absence* of transgenes in an area, landrace populations should be identified with the highest probability of containing them, *e.g.*, those in growing environments where TGVs and TGV × landrace hybrids would have highest fitness, and environments where farmers may have planted maize grain obtained from potential sources of TGV seed. If, on the other hand, the objective is determining the *frequency* of transgenes in the landrace populations of a given area, it is necessary to use a sampling methodology that maximizes the probability of finding rare alleles in the reference population (Crossa et al., 1993), and is representative of that population. For a given number of seeds collected it is important to take them from the maximum number of sampling units at each level, *e.g.*, ears and fields, and from as broad a range of environments within the reference population area as possible (Marshall and Brown, 1975). In addition, an equal number of seeds should be sampled from each sampling unit, beginning with the number of seeds/ear, number of ears/area, etc. (Vencovsky and Crossa, 1999).

The Quist and Chapela study (2001) was testing to see if transgenes were present in an area. They do not discuss their sampling strategy, but state that "As our samples originated from remote areas, it is to be expected that more accessible regions will be exposed to higher rates of introgression" (Quist and Chapela, 2001: 542). Because they found transgenes in a total sample of only six ears, it could be they followed some form of the targeted sampling strategy described above within the region selected, but no data were provided (2001).

Ortiz-García et al.'s sample was a very small proportion of fields and locations within their reference metapopulation, and the area sampled was not representative of Oaxaca (Tab. 1). They state that "maternal plants [ears] were considered to be independent because they were selected haphazardly" in fields, about 4/field in 2003, and 9/field in 2004 (calculated from Ortiz-García et al., 2005a), and in 2004 ears were taken from both "stressed" and "normal" plants. In 2004, Ortiz-García et al. sampled a total of 13 municipalities in the District of Ixtlán de Juárez out of a total of 26 in the District, and 68 in the Sierra Juárez Region, and 570 in the state of Oaxaca. Based on their map, most appear to be adjacent to the main highway through the district of Ixtlán de Juárez. In the following paragraphs we describe the Ortiz-García et al. sample in comparison to the sampling universe.

Localities. The total number of localities sampled was 16 each in 2003 and 2004, with two different (*i.e.* 18 total), or for 2004 0.15%, 2.46%, and 9.94% of localities at the state, region, and district level (calculations based on INEGI, 2004).

Fields. Census publications use "production units" (farming households), not fields, which are more numerous because farmers often have more than one field, *e.g.* in the Central Valleys and Isthmus mean fields/farmer = 3.2, maximum = 11 (Aragón Cuevas, Soleri and Cleveland, manuscript under preparation). A very conservative estimate, therefore, is that 1 production unit = 1 field. Ortiz-García et al. (2005a) sampled 81 fields in 2004, or 0.02%, 0.23%, and 1.14% of fields (production units) at the state, region, and district levels (calculations based on INEGI, 1996). The numbers of fields sampled in each locality were unequal (in 2004 range = 4–8 fields/locality, CV = 19.7%).

Ears. Based on studies in Oaxaca (Soleri et al., 2000) and Jalisco (Louette et al., 1997), and extensive research by one of the authors of this paper (FAC) in many areas of Oaxaca including the Sierra Juárez region, farmers nearly always select ears from a small proportion of their harvest, shelling only as many as necessary for sowing their fields. In the Sierra Juárez, farmers plant an average of 55058 seeds.ha^{-1}, with an average of 425 seeds/ear, equaling 130 ears.ha^{-1}, or 10% higher (143 ears.ha^{-1}) when seeds from the tip or butt are not used, a common farmer practice. Since on average 35–45000 plants.ha^{-1} are harvested, the proportion of independent maternal genotypes per field ≈0.003178 (143/45000). (The average field of 0.9 ha in the Sierra Juárez Region contains ~40500 plants at harvest.) This means it is critical to sample fields of as many farmers as possible. The numbers of ears sampled by Ortiz-García et al. (2005a) in each field were unequal (in 2004 range = 4–13 ears/field, CV = 21.0%).

Seeds. There appear to be no studies of paternity of individual seeds on single ears, but maize produces many millions of heavy pollen grains/plant (Baltazar and Schoper, 2002). Pollen dispersal is highly non-random, with a very high percentage of pollen deposited close to the plant (Jarosz et al., 2003; Ma et al., 2004; Raynor et al., 1972). Maize pollination biology is a reason for maximizing the number of farmers' fields and plants per field sampled. The number of seeds and ears sampled by Ortiz-García et al. (2005a) was variable between locations (in 2004 range = 117–206 seeds/ear, CV = 16.1%), but no data were provided for the number of seeds/ear within locations.

Table 1. Representativeness of field samples in tests for transgenes in Mexican maize landraces in Oaxaca, Mexico.

Sampling level	The sampling universe				Quist and Chapela	Ortiz-García et al.		
	State (Oaxaca)	Region/Rural District (Sierra Juárez)	District (Benemérito Distrito de Ixtlán de Juárez)	Municipality (Ixtlán de Juárez)		2003 sample	2004 sample	total
Regions (distrito rural)	7	.	.	.	1 (distrito rural Sierra Juárez)	1 (distrito rural Sierra Juárez)		
Districts (distritos)	30	3	.	.	1 (Benemérito distrito de Ixtlán de Juárez)	1 (Benemérito distrito de Ixtlán de Juárez)		
Municipalities (municipios)	570	68	26	.	1 (Ixtlán de Juárez)	13	13	14
Localities (localidades)	10519	650	161	15	2	16	16	18
Localities sampled by Ortiz-García et al. in 2004 (16)/total	0.0015	0.0246	0.0994					
Fields or production units (production units from INEGI 1996 Annuario Estadístico Oaxca: 427ff.)	368399	34480	7090	1106		43	81	124
Fields sampled by Ortiz-García et al. in 2004 (81)/prod units census	0.0002	0.0023	0.0114	0.0732				
Fields/locality	35.02	53.05	44.04	73.73		3	5	
Plants (ears)					6	164	765	929
Seeds						50126	103620	153746
Seeds/field						1166	1279	
Seeds/maternal plant		425			~150–400	306	147	
Seeds/locality						3133	6476	165

EFFECTIVE POPULATION SIZE

Changes in allele frequency caused by sampling error in small populations (random genetic drift) leads to a continuous fixation and loss of alleles, reducing the proportion of heterozygous individuals in the population (Crow and Kimura, 1970; Wright, 1931). These random changes in allele frequency that affect the genetic representativeness of the population are quantified and predicted using the parameter effective population size (N_e). Therefore, the census population size of individuals or seeds sampled (n) is not appropriate to use directly for estimating probabilities for detecting rare alleles (in this case transgenic alleles), because of variation in population size over time including due to drift, non-random contribution of individuals to the next generation, and other factors affecting the frequency of alleles.

N_e is the size of the ideal population (in Hardy-Weinberg equilibrium) showing the same amount of inbreeding (Hartl and Clark, 1989: 85) or allelic variance (Crossa and Vencovsky, 1994) as the census population, *i.e.* it is a measure of the genetic representativeness of a sample of individuals (Vencovsky and Crossa, 1999). The question is, therefore, what is the appropriate sampling scheme to be used to make N_e/n as large as possible? Crossa and Vencovsky (1994) and Vencovsky and Crossa (1999) have shown that the best strategy is to take equal numbers of seeds – a balanced sample – from the largest possible number of maternal plants within each sampling level.

We used the formula for variance effective population size (Hernandez and Crossa, 1993; Vencovsky and

Crossa, 1999),

$$N_{e(v)}=n/[(n–1)/4P+1],$$

where $_{(v)}$ indicates N_e based on allelic variance, n = census population size (number of seeds), P = number of parent plants contributing seeds or plants sampled. Because Ortiz-García et al. sampled a large number of seeds ($n = 103\,620$) from a small number of ears (765), the N_e for seeds (2972) for their sample is much smaller than n. Our estimate of N_e for the combined sample may even be conservative, because it does not include some important factors which could reduce it even further, such as the small number of ears that contribute to the plants in each field, discussed above.

Another important consideration is determining the level at which N_e should be calculated for a particular sample. This depends on knowledge of the genetic structure present within the reference population or metapopulation (Vencovsky and Crossa, 2003). For cases with a metapopulation comprised of multiple populations, Vencovsky and Crossa have shown that even when allelic diversity among populations (represented by F_{ST}) is small, sampling a small number of populations will reduce N_e and thus how representative the sample is of the metapopulation – "…it is impossible to measure N_e (for the metapopulation) appropriately if the structure of the metapopulation is unknown or, in other words, if the number of the component subpopulations… and the interpopulation diversity (F_{ST}) under real conditions are unknown" (Vencovsky and Crossa, 2003: 1915). That is, unless sampling can be based on specific knowledge of metapopulation genetic structure, N_e must be calculated for each population within the metapopulation and these N_e cannot be summed for a metapopulation estimate.

Quist and Chapela sampled 6 ears from 2 localities, but give no indication of ears per locality, though the possible range would be 1–5. No data are given on seeds/ear, but they state that each pooled sample of seeds from a single ear "represents a composite of ~150–400 pollination events" (Quist and Chapela, 2001: 541), inferring a maximum of 400 seeds/ear, which we use as a conservative estimate. The N_e for the total sample is then 24, and for localities ranges from 4 (1 ear/locality) to 20 (5 ears/locality).

As with Quist and Chapela's study, the metapopulation structure for the Ortiz-García study is unknown, that is the degree of divergence among the maize populations from the 16 localities sampled in 2004 was not evaluated, and therefore N_e must be calculated at the level of each locality separately, assuming there is no significant intralocality divergence. Calculation of each locality separately gives a range for N_e of 82–311.

Thus for each reference population (most appropriately localities in these studies), sampling should maximize the number of units (fields/locality, ears/field) sampled at each level, make consistent use of balanced samples (number of ears/field area, number of seeds/ear), and use N_e, not n, as the basis of any estimations of frequencies of rare alleles in the population.

STATISTICAL PROBABILITIES FOR ABSENCE OF EVIDENCE

It is necessary to base testing for transgene frequency in populations on population genetics principles including using N_e and not n. If n is used the results apply only to the probability of failing to detect a transgenic seed in the sample of seeds, give no indication of the frequency of transgenes in populations, and do not contribute to understanding the dynamics of transgene flow.

The probability of failing to detect a transgenic plant in landrace populations for Quist and Chapela's largest subsample (*i.e.* the locality with a possible 5 ears) ranges from 0.99960 for an expected frequency of 0.0001 to 0.817906 for an expected frequency of 0.1, suggesting that their sample was not random, or that transgene frequency was much greater than 0.1. (We calculated all binomial probabilities with SISA (Uitenbroek, 1997).)

With an alternative hypothesis that the true frequency of transgenic seeds was 0.0001, Ortiz-García et al. estimated that the "joint probability of failing to detect a single transgenic seed at any of the localities was… 0.00003 in 2004" based on the assumption that seeds they tested (103 620) were independent observations. They also assumed that detection of transgenes at the 0.0001 level in the two laboratories used was routine, and "with a degree of accuracy that is close to 100%" (Ortiz-García et al., 2005a), but no independent evaluation of this critical assumption was given (see also Paulson, 2005).

Lack of random sampling at the plant, field and location levels could mean that binomial probabilities are invalid. However, we will ignore these problems, and calculate binomial probabilities using Ortiz-García et al.'s data for seeds in 2004, for which their result was the lowest probability of not finding transgenes with an alternative hypothesis that they were present at a frequency of 0.0001. The range of probabilities for failing to detect transgenes in the landrace populations sampled in 2004 if present at 0.0001, using our estimates for N_e, are 0.96935–0.99183 for individual locations.

WHAT DO WE KNOW?

Based on the preceding analysis, we can now summarize what we know scientifically about maize transgenes in the samples taken in Oaxaca, Mexico using data from the two peer-reviewed papers published to date.

The Quist and Chapela sample was too small to use as the basis for estimating transgene frequency, and this was not their intent, although their results show transgenes were present in their very small sample of ears. We can estimate frequencies using Ortiz-García et al.'s sample, because it was much larger and from more localities. Table 2 shows the binomial probabilities of failing to detect transgenes at expected frequencies from 0.0001 to 0.1. The probability of failing to detect a transgene becomes significant ($P < 0.05$) at transgene frequencies between 0.01 and 0.1 for the locality with the largest and smallest N_e respectively. Another way of expressing this is that at a binomial probability of 0.05, the upper limit of expected frequency of transgenes in the populations sampled by Ortiz-García et al. that would fail to be detected ranges between 0.00961 and 0.03586 (*i.e.* ~1–4%). Based on an estimate by one of the authors (FAC) of 17000 ha planted to maize landraces in the Sierra Juárez Region and 6233 in the Ixtlán de Juárez District, and a conservative estimate of 35000 plants.ha^{-1} harvested, Ortiz-García et al.'s sample could have failed to detect >5.7 million transgenic maize plants in the Region, or >2.0 million within the District. Thus, Ortiz-García et al.'s analysis shows that transgenes were probably not present in the sampled communities at frequencies greater than 1–4%, but that many transgenic plants could be present but undetected.

In terms of Oaxaca, the two papers tell us very little since the Sierra Juárez region is not typical of that state, where ~90% of total maize area is planted with landraces. The Sierra Juárez region is a mountainous area, and only 10% of Oaxacan maize landrace area is in the mountains (3.5% in the Sierra Juárez and 6.5% in the Sierra Sur) (calculated from Aragón Cuevas et al., 2005). Forestry, not maize cultivation, is the primary activity of the rural population there. The Ixtlán de Juárez district contains only ~21% (7090/34480) of the fields in the Region. This is not the best area for detecting transgene presence in Oaxacan landraces – transgene flow to landraces is likely to be lower here compared with large, contiguous areas of maize that include some less stressful environments, *e.g.*, the Central Valleys of Oaxaca (25% of the state's maize landrace area). In the central valleys, maize is cultivated year round under both rainfed and irrigated conditions.

SUGGESTED METHODS FOR DETECTING TRANSGENES IN LANDRACE POPULATIONS

More research with explicit methodologies and documented assumptions is needed as the basis for policy decisions about transgene flow and transgenic crop varieties in centers of crop origin and/or diversity, such as maize in Mexico. We suggest that this research be based on (1) understanding local seed systems, including the way farmer practices affect landrace population structure and dynamics, (2) collecting seed samples to optimize N_e/n, generally by taking a smaller number of seeds from a larger number of ears than has been done, (3) minimizing variance by sampling equal numbers of units at all levels, and (4) using N_e, not n, to calculate binomial probabilities for presence of transgenes. The sampling methods used by Quist and Chapela and Ortiz-García et al. are commonly used in the collection of germplasm and for research because of the reality of limited time and resources. However, the controversy surrounding transgene flow to landraces, especially maize in Mexico, and the need for policy to be based on sound science, makes this case different, requiring a more robust sampling strategy and a more cautious interpretation of results.

Assuming the ideal condition in which $n = N_e$, we calculated the n that would be required to detect at frequencies of 0.0001 to 0.01, with 0.95 probability, at least one copy of a transgenic allele, when there is one transgenic locus and either one allele per locus (the hemizygous condition characterizing the initial generations of introgression) or two alleles (homozygous condition characterizing the end point of introgression), reflecting the fact that at any one time there may be gene flow/introgression cases of different ages and states of advancement. Because $n = N_e$ only when each seed comes from a different independent plant, the goal is to obtain an N_e of the same size as n identified in the calculations. Thus, depending on the transgene frequency, the N_e required is ~300–60000 (Tab. 3). For example, to detect transgenes in a population in a hemizygous condition at a frequency of 0.001, an $N_{e(v)}$ of ~6000 is required. This could be obtained by analyzing 400 seeds/plant for 1500 plants (ears) for a total of 600000 seeds, 50 seeds/plant for 1600 plants (ears) for a total of 80000 seeds, or 2 seeds/plant for 4500 plants (ears) for a total of 9000 seeds. The combination of seeds/ear and number of ears will depend on balancing the costs of collecting samples in the field and analyzing seeds in the lab, but in general it is likely that a high N_e/n (smaller number of seeds/ear, larger number of ears) will be the most efficient, especially when the subpopulation structure is unknown, which will almost always be the case when

Table 2. Variance effective population size and binomial probabilities of failure to detect transgenes at different frequencies, for data of Ortiz-Garcia et al. (2005) for 2004. Variance effective population size calculated using $N_{e(v)}=n/[(n-1)/4P+1]$ (Vencovsky and Crossa, 1999).

Localities sampled by Ortiz-Garcia et al.	Number of seeds in 2004 analyzed by Genetic ID	analyzed by GeneScan	Plants (=ears) seeds taken from, 2004 (P)	total (n)	$N_{e(v)}$	For calculated $N_{e(v)}$, binomial probability of failing to detect a transgene, with 95% CI, if present at a frequency of 0.0001	0.001	0.01	0.1	For calculated $N_{e(v)}$, detection level frequency of transgene at binomial probability = 0.05, 95% CI	No. of transgenic plants possible at a given detection level frequency when binomial prob = 0.05 Distrito Rural Sierra Juárez	Benemérito Distrito de Ixtlán de Juárez
Locality (Santiago Comaltepec) with smallest $N_{e(v)}$	2160	2160	21	4320	82	0.99183	0.92085	0.43862	0.00017	0.03586	21 335 602	7 822 636
Locality (San Francisco la Reforma [San Pedro Yolox]) with largest $N_{e(v)}$	5630	5630	80	11260	311	0.96935	0.73247	0.04383	0.00000	0.00961	5 719 757	2 097 132

Table 3. n (census population size) required to obtain at least one copy of a transgenic allele with one locus, and one or two alleles, if present in the population at different frequencies, $\alpha = 0.05$ [P=(1–α)=0.95]. Assumes ideal condition of $n = N_{e(v)}$, else goal for $N_{e(v)}$ is n given here.

Transgene frequency	n (homozygous, *i.e.* two alleles)	n (hemizygous, *i.e.* one allele)
0.0001	29 996	59 992
0.001	2996	5992
0.01	298	596

sampling landraces in the field. The sampling scheme at the region, district, locality, field and ear level needs to minimize variance, including by taking equal numbers of seeds from ears, ears from fields, etc, and sampling units representative of the landrace populations for which results are intended to be extrapolated (*e.g.* the State of Oaxaca, see above).

Sampling could be based on a combination of methods to both (a) detect any transgenes if present, and to (b) determine frequencies of transgenes. Because very large numbers of ears would be necessary to detect transgenes in populations at low frequencies, sample areas could first be selected where the likelihood of transgene presence is highest. In Oaxaca this means areas like the central valleys, communities with irrigation where maize is grown year-round making it more likely that a transgenic variety could grow, and where Diconsa stores have been active over the last 10 years. Random samples of farmers in these communities could be interviewed to identify those who have planted seeds obtained from Diconsa, and plants from their fields sampled. Those farmers' fields, and perhaps fields adjacent to theirs, could be targeted for samples to test for transgene presence. A random sample, stratified by important distinguishing variables identified in the interviews (*e.g.*, seed system characteristics, area sown to maize, wealth, etc.), would be more appropriate for determining transgene frequency in that community. In both cases, taking a small number of seeds from a large number of ears, equally distributed among those farmers maximizes N_e/n. Maize ears are stored in the husk to reduce insect damage, and families we have interviewed shell about five ears for eating per person daily. Therefore, a goal is to have enough families in the sample that the required number of ears could be sampled from those shelled daily over a period of 1–3 days. It would be important to sample soon after harvest to minimize under representing transgenic ears, as these will likely be most prone to insect damage in storage, as has been observed by Oaxacan farmers and researchers of MVs and FV $\times$ MV

hybrids, and documented in comparisons of some hybrids and Oaxacan landrace accessions for resistance to maize weevil (*Sitophilus zeamais*) (Savidan, 2002).

As a control, samples should also be made in communities where transgenes are least likely, *e.g.* remote communities and those with particularly difficult growing conditions.

CONCLUSION

Transgene flow has potentially important genetic, economic, social and cultural effects. The importance of detecting unintentional presence of transgenes in landraces in centers of origin and/or crop genetic diversity will increase with an increasing variety of commercially approved TGVs. This is especially true for the new generation of TGVs engineered to produce pharmaceutical and industrial chemicals being developed in such food crops as maize and rice.

We agree with Quist and Chapela that "Further study of the impact of the gene flow from commercial hybrids to traditional landraces in the centers of origin and diversity of crop plants needs to be carefully considered" (Quist and Chapela, 2001: 542), and with Ortiz-García et al. that "Worldwide, the ability of transgenes to disperse among countries merits closer scientific monitoring, especially when the receiving country has not approved them for environmental release, and in cases when the genetic modification eliminates or impairs the use of a particular crop as food" (Ortiz-García et al., 2005a: 12343).

It is possible that the *data* of both Quist and Chapela and Ortiz-García et al. could accurately describe the situation today in the locations they sampled. However, due to the sampling issues discussed, how those data are *interpreted* must be carefully evaluated. While both studies have sampling problems, these are more serious when the null hypothesis is accepted, as Ortiz-García et al. did, because type II error is critical. If the evidence presented by Quist and Chapela and other researchers of transgene presence in maize landraces is accepted, Ortiz-García et al.'s data are valuable for establishing that transgenes are not present at a frequency greater than 1–4% in the sampled localities, but do not show that they have been reduced to extremely low frequencies, or have disappeared. If evidence suggesting transgene presence is not accepted, we still do not have any data to support the proposition that transgenes are not present at other localities, or at frequencies below 1–4% in the localities in the Ortiz- García et al. study.

While it is likely that transgenes are present today in maize landraces in Mexico, we have no scientific evidence

of their presence or absence for the last several years in the State of Oaxaca, or in Sierra Juárez region. Thus, at this time, there appears to be no justification for making inferences (as have *e.g.* Kaiser, 2005; Marris, 2005; Ortiz-García et al., 2005a) about the frequency of maize transgenes outside of the communities sampled in the Ixtlán de Juárez district of the Sierra Juárez region or about the possible effects of transgene presence on the maize populations sampled or the communities that manage them. This means that the data of Ortiz-García et al. provide no basis for Mexican or Oaxacan policy regarding maize transgene flow, as has been assumed by some (Prakash, 2005; Raven, 2005; Velasco, 2005), except that more research is required.

As Quist and Chapela (2001), Ortiz-García et al. (2005a) and others have pointed out, the detection of presence or absence of transgenes, and their frequencies, in landrace populations at different points in space and time is an important tool for understanding transgene flow to landrace and wild and weedy crop relative populations, and its potential effects. However, statements about the presence or absence of transgenes in these populations need to be based on sound scientific method and theory, especially if used as the basis for policy. Scientifically unjustified conclusions contribute to misunderstanding – and may lead either to false alarm or false complacency.

ACKNOWLEDGEMENTS

REFERENCES

Altman D, Bland JM (1995) Absence of evidence is not evidence of absence. *Brit. Med. J.* **311**: 485

Alvarez-Morales A (2000) Mexico: Ensuring environmental safety while benefiting from biotechnology. In Persley GJ and Lantin MM, eds, Agricultural biotechnology and the poor: proceedings of an international conference, Washington, D.C., 21-22 October 1999. Consultative Group on International Agricultural Research, Washington, DC

Alvarez-Morales A (2002) Transgenes in maize landraces in Oaxaca: official report on the extent and implications. http://www.bba.de/gentech/isbgmo.pdf (posted 2002 Oct 10-16; verified 2006 Jan 24)

Andow DA (2003) Negative and positive data, statistical power, and confidence intervals. *Environ. Biosafety Res.* **2**: 1–6

Aragón Cuevas F, Taba S, Castro-García FH, Hernández-Casillas JM, Cabrera-Toledo JM, Alcalá LO, Ramírez ND (2005) In situ conservation and use of local maize races in Oaxaca, Mexico: a participatory and decentralized approach. In Taba S ed, Latin American maize germplasm conservation: regeneration, in situ conservation, core subsets, and prebreeding; proceedings of a workshop held at CIMMYT, April 7-10, 2003. CIMMYT, Mexico

Ballinas V, Becerril A (2005) Aprueba el senado la ley de bioseguirdad, pese a deficiencias, La Jornada, México

Baltazar BM, Sanchez-Gonzalez JD, de la Cruz-Larios L, Schoper JB (2005) Pollination between maize and teosinte: an important determinant of gene flow in Mexico. *Theor. Appl. Genet.* **110**: 519–526

Baltazar BM, Schoper JB (2002) Crop-to-crop gene flow: dispersal of transgenes in maize during field tests and commercialization. Paper presented at the 7th International Symposium on the Biosafety of Genetically Modified Organisms, Oct 10-16, 2002, Beijing, China. http://www.worldbiosafety.net/ (posted 2002 Oct 10-16; verified 2006 Jan 24)

Berthaud J, Gepts P (2004) Chapter three: Assessment of effects on genetic diversity. In: Maize and biodiversity: The effects of transgenic maize in Mexico. Commission for Environmental Cooperation of North America http://www.cec.org/files/PDF//Maize-Biodiversity-Chapter3_en.pdf (verified 2006 Jan 7)

CEC (2004) Maize and biodiversity: the effects of transgenic maize in Mexico. Key findings and recommendations. Secretariat article 13 report. CEC. http://www.cec.org/files/PDF//Maize-and-Biodiversity_en.pdf (verified 2004 Nov 10)

Cleveland DA, Soleri D (2005) Rethinking the risk management process for GE crops in Third World agriculture. *Ecol. Soc.* **10**: Article 9. http://www.ecologyandsociety.org/vol10/iss1/art9/

Crossa J, Vencovsky R (1994) Implication of the variance in effective population size on the genetic conservation of monoecious species. *Theor. Appl. Genet.* **89**: 936–942

Crossa J, Hernandez CM, Bretting P, Eberhart SA, Taba S (1993) Statistical genetic considerations for maintaining germ plasm collections. *Theor. Appl. Genet.* **86**: 673–678

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper and Row, New York

Ellstrand NC (2003) Dangerous liaisons? When cultivated plants mate with their wild relatives. Johns Hopkins University Press, Baltimore

Esteva G, Marielle C, eds (2003) Sin maíz no hay país. Consejo Nacional para la Cultura y las Artes, Dirección General de Culturas Populares e Indígenas, México

Gepts P, Papa R (2003) Possible effects of (trans)gene flow from crops on the genetic diversity from landraces and wild relatives. *Environ. Biosafety Res.* **2**: 89–103

Hall L, Topinka K, Huffman J, Davis L (2000) Pollen flow between herbicide-resistant *Brassica napus* is the cause of multiple-resistant *B. napus* volunteers. *Weed Sci.* **48**: 688–694

Hartl DL, Clark AG (1989) Principles of population genetics. Sinauer Associates, Sunderland, Massachusetts

Hernandez CM, Crossa J (1993) A program for estimating the optimum sample-size for germplasm conservation. *J. Hered.* **84**: 85–86

Huang JK, Hu RF, Rozelle S, Pray C (2005) Insect-resistant GM rice in farmers' fields: Assessing productivity and health effects in China. *Science* **308**: 688–690

INEGI (1996) Anuario Estadístico del Estado de Oaxaca, Edición 1996. INEGI, Aguascalientes, Mexico

INEGI (2004) *Sistema para la consulta del anuario estadístico de Oaxaca, Edicion 2004*. http://www.inegi.gob.mx/est/contenidos/espanol/sistemas/aee04/estatal/oax/index.htm (verified 2005 Oct 8)

Jarosz N, Loubet B, Durand B, McCartney A, Foueillassar X, Huber L (2003) Field measurements of airborne concentration and deposition rate of maize pollen. *Agr. Forest Meteorol.* **119**: 37–51

Kaiser J (2005) Calming fears, no foreign genes found in Mexico's maize. *Science* **309**: 1000

Kaplinsky N, Braun D, Lisch D, Hay A, Hake S, Freeling M (2002) Maize transgenic results in Mexico are artefacts. *Nature* **416**: 601

Louette D, Charrier A, Berthaud J (1997) In situ conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. *Econ. Bot.* **51**: 20–38

Lu BR, Snow AA (2005) Gene flow from genetically modified rice and its environmental consequences. *Bioscience* **55**: 669–678

Ma BL, Subedi KD, Reid LM (2004) Extent of cross-fertilization in maize by pollen from neighboring transgenic hybrids. *Crop Sci.* **44**: 1273–1282

Marris E (2005) Four years on, no transgenes in Mexican maize. *Nature* **436**: 760

Marshall DR, Brown AHD (1975) Optimum sampling strategies in genetic conservation. In Frankel OH and Hawkes JG eds, Crop genetic resources for today and tomorrow, Cambridge University Press, Cambridge, U.K., pp 53–80

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **99**: 6080–6084

Metz M, Fütterer J (2002) Suspect evidence of transgenic contamination. *Nature* **416**: 600–601

Nature Biotechnology (2002) Editorial: Going with the flow. *Nat. Biotechnol.* **20**: 527

NRC (2002) Environmental effects of transgenic plants: the scope and adequacy of regulation. National Academy Press, Washington, DC

NRC (2004) Biological confinement of genetically engineered organisms, Washington, DC

Ortiz-García S, Ezcurra E, Schoel B, Acevedo F, Soberón J, Snow AA (2005a) Absence of detectable transgenes in local landraces of maize in Oaxaca, Mexico (2003-2004). *Proc. Natl. Acad. Sci. USA* **102**: 12338–12343

Ortiz-García S, Ezcurra E, Schoel B, Acevedo F, Soberón J, Snow AA (2005b) Correction. *Proc. Natl. Acad. Sci. USA* **102**: 18242

Papa R, Acosta J, Delgado-Salinas A, Gepts P (2006) A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theor. Appl. Genet.* **111**: 1147–1158

Paulson KM (2005) Found-and-lost: transgenic maize in Oaxaca, Mexico. Information Systems for Biotechnology http://www.isb.vt.edu/news/2005/news05.nov.htm#nov0502 (verified 2005 Nov 07)

Peregrina K, Crúz J (2005) Mexico approves planting and sale of GM crops. Source: SciDev.Net. http://www.scidev.net/News/index.cfm?fuseaction=readNews&itemid=1945&language=1 (verified 2006 Jan 24)

Prakash CS (2005) Duh.... No GM genes in Mexican corn. AgBioWorld. http://www.agbioworld.org/newsletter_wm/index.php?caseid=archive&newsid=2398 (verified 2005 Aug 25)

Pressoir G, Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* **92**: 95–101

Quist D, Chapela IH (2001) Transgenic DNA introgressed into traditional maize landraces in Oaxaca, Mexico. *Nature* **414**: 541–543

Quist D, Chapela IH (2002) Quist and Chapela reply. *Nature* **416**: 602

Raven PH (2005) Transgenes in Mexican maize: desirability or inevitability? *Proc. Natl. Acad. Sci.* **102**: 13003–13004

Raynor GS, Ogden EC, Hayes JV (1972) Dispersion and deposition of corn pollen from experimental sources. *Agron. J.* **64**: 420

Savidan A (2002) Tritrophic interactions in maize storage systems. Institute of Zoology, Laboratory of Animal Ecology and Entymology, University of Neuchâtel, Neuchâtel, p 225

Serratos JA, Willcox MC, Castillo-González F, eds (1997) Gene flow among maize landraces, improved maize varieties and teosinte: implications for transgenic maize. CIMMYT, Mexico

Snow AA, Andersen B, Jorgensen RB (1999) Costs of transgenic herbicide resistance introgressed from Brassica napus into weedy B-rapa. *Mol. Ecol.* **8**: 605–615

Snow AA, Pilson D, Rieseberg LH, Paulsen MJ, Pleskac N, Reagon MR, Wolf DE, Selbo SM (2003) A *Bt* transgene reduces herbivory and enhances fecundity in wild sunflowers. *Ecol. Appl.* **13**: 279–286

Soleri D, Smith SE, Cleveland DA (2000) Evaluating the potential for farmer and plant breeder collaboration: a case study of farmer maize selection in Oaxaca, Mexico. *Euphytica* **116**: 41–57

Soleri D, Cleveland DA, Aragón Cuevas F, Ríos Labrada H, Fuentes Lopez MR, Sweeney SH (2005) Understanding the potential impact of transgenic crops in traditional agriculture: maize farmers' perspectives in Cuba, Guatemala & Mexico. *Environ. Biosafety Res.* **4**: 141–166

Soleri D, Cleveland DA, Aragón Cuevas F (2006) Transgenic crop varieties and varietal diversity in traditionally based agriculture: the case of maize in Mexico. *Bioscience* **56**: 503–513

Uitenbroek DG (1997) "*SISA-Binomial*". http://home.clara.net/sisa/binomial.htm (verified 2006 Jan 09)

Velasco CE (2005) El maíz criollo de Oaxaca, libre de contaminación genética: cientificos, *La Jornada*, México

Vencovsky R, Crossa J (1999) Variance effective population size under mixed self and random mating with applications to genetic conservation of species. *Crop Sci.* **39**: 1282–1294

Vencovsky R, Crossa J (2003) Measurements of representativeness used in genetic resources conservation and plant breeding. *Crop Sci.* **43**: 1912–1921

Wright S (1931) Evolution in Mendelian populations. *Genetics* **16**: 97–159

Zi X (2005) GM rice forges ahead in China amid concerns over illegal planting. *Nat. Biotechnol.* **23**: 637

To access this journal online:
www.edpsciences.org