Transgenic Res (2014) 23:135–143
DOI 10.1007/s11248-013-9737-0

ORIGINAL PAPER

No effects of *Bacillus thuringiensis* maize on nontarget organisms in the field in southern Europe: a meta-analysis of 26 arthropod taxa

C. Comas · B. Lumbierres · X. Pons · R. Albajes

Received: 28 February 2013 / Accepted: 23 July 2013 / Published online: 2 August 2013

Abstract Maize with the insecticidal properties of the entomopathogenic bacterium *Bacillus thuringiensis* Berliner, known as Bt maize, has been sown in Europe since 1998. For several years, EU and Spanish regulations have required laboratory and field trials to assess risks of genetically modified crops for nontarget organisms prior to their authorization. Thirteen field trials were conducted in Spain to measure the effects of Bt maize on a broad range of arthropod taxa; no effects were found in accordance with most literature records. However, statistical analyses of single trials rarely have the statistical power to detect low effect sizes if they do not have a sufficient sample size. When sample size is low, meta-analysis may improve statistical power by combining several trials and assuming a common measure of effect size. Here we perform a meta-analysis of the results of 13 independent field trials conducted in Spain in which effects of single or stacked Bt traits on several arthropod taxa were measured with no significant results. Since the taxa included in each single trial were not the same for all trials, for the meta-analysis we selected only those taxa recorded in a minimum of six trials, resulting finally in 7, 7, and 12 taxa analyzed in visual counts, pitfall traps and yellow sticky traps, respectively. In comparison with single trial analysis, meta-analysis dramatically increased the detectability of treatment effects for most of the taxa regardless of the sampling technique; of the 26 taxa analyzed, only three showed poorer detectability in the meta-analysis than the best recorded in the 13 single trials. This finding reinforces the conclusion that Bt maize has no effect on the most common herbivore, predatory and parasitoid arthropods found in the maize ecosystems of southern Europe.

Keywords Meta-analysis · Nontarget arthropods · NTO · GM corn · Bt

C. Comas (✉)
Department of Mathematics, Agrotecnio Center, Universitat de Lleida, Catalonia, Spain
e-mail: carles.comas@matematica.udl.cat

B. Lumbierres · X. Pons · R. Albajes
Department of Crop and Forest Sciences, Agrotecnio Center, Universitat de Lleida, Catalonia, Spain

Introduction

Maize (*Zea mays* L.) has been grown on over 10 million ha in Europe in recent years (Czarnak and Rodríguez-Cerezo 2010). Among the insects affecting maize yield, the Lepidopteran borers [*Ostrinia nubilalis* (Hübner) (Lepidoptera: Crambidae) and *Sesamia nonagrioides* Lefèbvre (Lepidoptera: Noctuidae)] are the main pests (Meissle et al. 2010). In addition to biological control and cultural practices, genetically modified (GM) varieties with the insecticidal properties of the entomopathogenic bacterium *Bacillus thuringiensis* Berliner (Bt maize) may provide

efficient control of maize borers and reduce chemical applications (Meissle et al. 2010). A survey conducted among growers in Spain estimated that Bt maize is economically very efficient in areas with a high pest pressure and may replace most of the insecticidal sprayings carried out to control these pests (Gómez-Barbero et al. 2008). Spain is the main grower of Bt maize in Europe; from its authorization for cultivation in 1998–2012 the area of Bt maize has increased to 116,030 ha, representing 30 % of the total area under maize in the country (MAGRAMA 2012). In areas where maize borers are particularly damaging, such as the study area in Aragon and Catalonia (NE Iberian Peninsula), the concentration of Bt maize may reach 70 % of the maize grown for grain.

Bt maize has proven to be effective for controlling borers and has a potential role in containing expansion of *Diabrotica virgifera virgifera* LeConte (Coleoptera: Chrsyomelidae), an important pest in the USA that was introduced into Europe in the early 1990s. Furthermore, Bt maize has environmental benefits because it reduces the need for chemical applications. However its deployment has prompted extensive debate over risks for the environment and particularly for nontarget organisms, including arthropods. Arthropods provide important ecological services in maize ecosystems such as biological control, pollination and decomposition, and they form an important part of the biodiversity, so direct or indirect effects of Bt maize on arthropods may interfere with these services. The mandate to conduct laboratory and field trials to assess risks of GM crops for the environment and particularly for nontarget organisms prior to their authorization has been in place in the regulation of the EU and member States for several years (EFSA 2010). In response to this mandate, governmental bodies and companies have promoted and sponsored many European Research Area (ERA) activities in Europe to measure potential effects of GM maize on nontarget arthropods (NTAs) in both the laboratory and the field. A tiered approach to testing potential side effects of GM crops on NTAs, from the laboratory to the field, was proposed by Romeis et al. (2008). Usually, it is necessary to conduct field trials or use models when a GM crop has proven harmful in the laboratory or when the main potential effects are the result of complex interactions among many factors that cannot be studied in simple laboratory conditions. However, this field testing is difficult to interpret due to the interaction of many factors that cannot be studied in simple laboratory conditions.

In the last 12 years field trials for ERA purposes have been conducted in several European countries, including Spain, where the authors have assessed risks of GM Bt and herbicide-tolerant (HT) maize varieties for NTAs (Comas et al. 2013). Only occasional effects of Bt maize varieties on herbivore, predatory or parasitoid insects have been reported (Lumbierres et al. 2004, 2011; Pons et al. 2005; de la Poza et al. 2005; Albajes et al. 2012), and many authors have concluded that the involvement of the Bt traits in these occasional effects has not been proven (see the review of Naranjo 2009). However, the lack of negative effects could be due to the insufficient statistical power of the tests. The power of a test is the probability of rejecting the null hypothesis, no effects, when it is false and the alternative hypothesis is true. Then it measures the probability that the test detects an effect of a known magnitude using a specified experimental design and varies according to the magnitude of the effect specified (Perry et al. 2003). Inversely, the magnitude of an effect (the effect size) that a test is able to detect may be calculated for a specified power. The effect size is usually obtained as the scaled difference between the density of the organism recorded on the GM variety and the density of the organism on a control non-GM variety called the comparator. This value can be expressed as a percentage of the control mean density. Once the power of the test has been specified, we can obtain the minimum difference between the density of the organism on the control non-GM and the GM variety, given as a percentage of the control mean density, that the test is able to detect. This minimum difference is the detectability of the test. A field test with a high detection capability, for a given test power, is a test that is able to detect small significant difference between the control non-GM and the GM variety. Therefore, an improvement in the detectability of a test implies better detection of small significant effects.

Increasing the statistical power of individual field tests to satisfactory values would involve increasing sample size by increasing the number of replications, treatments or years/sites of trials, a rather costly approach. Alternatively, if several trials are available, a meta-analysis may improve statistical power by combining them and assuming a common measure of effect size. In fact, this approach can be used to integrate several independent trials, whether published

or unpublished, that were not initially defined to be combined and thus obtain new and more robust results (Borenstein et al. 2009). This approach was used by Marvier et al. (2007), Wolfenbarger et al. (2008) and Naranjo (2009) to study the nontarget effects of Bt crops reported in 42, 45 and 63 field studies, respectively.

This study aimed to determine whether the no-effect conclusions reached by ANOVA analysis of single field trials with Bt maize varieties conducted in Spain from 2000 to 2010 is confirmed by a meta-analysis of all these trials. Complimentarily, the effect detection capacity of taxa recorded in field trials is calculated with a meta-analysis approach and compared with values of single trial analysis. To this end, among the 20 field trials carried out by authors in the period to measure nontarget effects of GM maize (Bt, HT and stacked traits), we selected 13 trials in which Bt and near-isogenic non-Bt varieties were compared for population density of several taxa, including arthropod herbivores, predators, omnivores, parasitoids and decomposers.

Materials and methods

Field trials and arthropod sampling

From 2000 to 2010, the authors carried out 20 field trials to test the effect of transgenic traits introduced into maize on several NTAs. The traits studied were the insecticidal capacity of Bt (several single or stacked Bt Cry toxins), the tolerance to broad spectrum herbicides (HT) and several Bt and HT stacked traits. The trials covered a range of characteristics that are common in trials performed to measure nontarget effects of GM crops: different numbers of sampling dates (4–8), numbers of treatments (2–10 with transgenic vs. near-isogenic varieties, pesticide vs. pesticide-free treatments, and a certain number of reference varieties), and different numbers of single Bt or HT, or stacked traits (1–3). Only trials with single or stacked Bt traits and near-isogenic controls (a total of 13 trials) were used for the analysis, The trials were conducted from May to early October at several locations of the Lleida region (NE Iberian Peninsula), an area where more than 70 % of the corn grown in recent years has been Bt because of the high pressure of corn borer populations. The area has a Mediterranean climate with high temperatures (max T between 25 and 40 °C) and a mean rainfall lower than 200 mm during the season, so corn may be grown only with irrigation. The experimental corn plots had an area of between 1,000 and 5,000 m^2 and they were always arranged in a randomized complete block design, with 3 or 4 blocks. NTA abundance or activity was estimated by visual counting, capture in pitfall and yellow sticky traps the sampling techniques most used for nontarget studies (see Albajes et al. 2009 for details). Arthropods were identified to species, genus, family or order level.

We aimed to test the hypothesis that Bt varieties affect the abundance of NTAs. For comparisons between the two kinds of varieties, we used the seasonal mean abundance (average over multiple sampling dates), as performed in Comas et al. (2013).

The meta-analysis

We performed a meta-analysis for each taxon and for each sampling technique. For each meta-analysis, we used the raw difference between the means of Bt and the control treatments (the near-isogenic variety) as a measure of effect size. Negatives values indicate lower taxa abundance in Bt plots compared with control ones and, inversely, positive values indicate higher abundance on Bt plots. We considered the raw mean difference instead of the standardized mean difference (Hedge's d) because all the studies considered in a given meta-analysis are based on the same scale of magnitude, so it is not necessary to provide a dimensionless index to combine them. Notice that the resulting meta-analyses depend on the magnitude scale of both taxa and sampling technique, though direct comparisons between meta-analyses are not performed in this study. Moreover, the use of the raw difference between means allows us to consider the detectability of each meta-analysis as defined above, and to compare our meta-analysis results with those obtained for single studies by Comas et al. (2013). Also note that the use of different sample sizes for each study (3–4 sample sizes) does not affect the resulting meta-analysis when the raw difference is used,

because this sample size effect is reflected and incorporated in the variance associated with this raw difference for each study and a given meta-analysis.

The trials considered in each meta-analysis are assumed to be independent from each other and they are defined by the year of the field trial and the trial location. Taxa with fewer than six trials per meta-analysis were not included in the analysis. Table 1 shows the number of trials for each meta-analysis performed in terms of taxa and sampling technique. We used fixed effect meta-analyses (Borenstein et al. 2009; Whitehead 2002) because all the trials, regardless of the taxa and the sampling technique, were functionally identical (i.e. they had all been carried out by the same research group with the same study design), so we expected to estimate a common effect size for each meta-analysis. We also computed a measure of heterogeneity for each meta-analysis to assess the validity of our assumption (Borenstein et al. 2009). In particular, for each meta-analysis we computed the statistic Q (a Q test) (Whitehead 2002). As the statistic Q follows a χ^2 distribution with $n-1$ degrees of freedom, a test of heterogeneity can be performed by comparing the theoretical mean value of Q under a χ^2_{n-1} distribution with the empirical one.

In 3 of 26 meta-analyses we found significant differences between the expected and the empirical Q, so we rejected the null hypothesis that all the trials have a common effect. For these 3 cases we assumed random effect meta-analyses to account for this heterogeneity (Borenstein et al. 2009).

A measure of detectability of treatment effects based on meta-analysis

To measure the detectability of treatment effects for our meta-analyses, we used the raw difference between means of Bt and the control treatments (i.e. the estimated common effect for a given meta-analysis), divided by the mean of the control treatment for this meta-analysis. Absolute small values of this measure indicate a high degree of detectability of small abundance changes between Bt and control treatments, whereas absolute large values suggest a poor detectability of taxon abundance between the two treatments, for a given meta-analysis. In particular, we obtained the detectability measure through the power of the test formula for a given meta-analysis approach. The power formula for a two-tailed test for a fixed effect meta-analysis is defined as (Borenstein et al. 2009)

$$Power_1 = 1 - \Phi(z_{1-\alpha/2} - \lambda) + \Phi(-z_{1-\alpha/2} - \lambda) \quad (1)$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution, $z_{1-\alpha/2}$ is the standard normal deviate for α level of probability (i.e. the Type I error), and

$$\lambda = \frac{Y}{\sqrt{Var[Y]}} \quad (2)$$

where Y is the (hypothesized) *true* common effect size for the meta-analysis approach. Now the detectability of the meta-analysis approach is

$$d_m = \frac{Y}{E[m_c]} 100\,\% \quad (3)$$

where $E[m_c]$ is the expected value of the abundance for the control treatment. Once the two errors had been set (i.e. the Type I (α) and II (β) errors being the power $1-\beta$), we obtained from (1) the value of λ. In practice, λ and d_m depend on $Var[Y]$ and $E[m_c]$, respectively, which are unknown and must be estimated. Then, we substitute $Var[Y]$ by $Var[\hat{Y}]$, i.e. the variance of the estimated common effect size for the meta-analysis variance and $E[m_c]$ can be estimated by $\bar{m}_c$, the average abundance of the control treatment for the n trials considered in the meta-analysis. Note that Expression (1) can only be solved numerically for a given power and a $z_{1-\alpha/2}$ value. If we assume the standard practice to set $\alpha = 0.05$ and a power of 0.8 (i.e. a Type II error $\beta = 0.2$), a numerical solution of (1) gives $\lambda = 2.80158$. Then, from (2) and (3)

$$d_m = \frac{2.80158\sqrt{Var[\hat{Y}]}}{\bar{m}_c} 100\,\%. \quad (4)$$

Similarly, for a random effect meta-analysis the detectability measure can be written as

$$d^*_m = \frac{2.80158\sqrt{Var^*[\hat{Y}]}}{\bar{m}_c} 100\,\% \quad (5)$$

where $Var^*[\hat{Y}]$ is the variance of the size effect of a random effect meta-analysis (i.e. the within and between study variability) (see Borenstein et al. 2009). We computed this detectability measure for individual trials to compare the effect of assuming a meta-analysis approach. In this case, we used a two-tailed t test to assess the null hypothesis that, for a

Table 1 Number of studies for each meta-analysis in terms of taxa and sampling technique together with the relative effect sizes (effect sizes divided by the average of the abundance for the control), detectability measure Eqs. (4) and (5) (asterisk values) and the corresponding maximum and minimum study detectability measured in single trials Eq. (8) for a given meta-analysis and $\alpha = 0.05$ and power 0.8. Max (%) and Min (%) stand for the maximum and minimum detectability values for each meta-analysis

Sampling technique	Taxa	No. of studies	Rel. effect size	Detectability (%)	Max (%)	Min (%)
Visual	*Orius* spp.	13	0.0253	6.67	66.48	11.56
	Nabis spp.	13	0.068	16.82	216.82	29.15
	Carabidae	11	0.0429	38.11	528.89	60.71
	Chrysopidae	11	0.0377	8.64	331.06	28.29
	Coccinellidae	11	−0.0228	6.41	731.82	26.61
	Araneae	12	0.0092	3.62	336.68	10.83
	Total predators	13	0.0243	6.24	45.01	9.05
Pitfall	Dermaptera	13	0.0082	8.14	267.39	40.05
	Carabidae	13	0.0374	7.77	91.37	10.89
	Staphylinidae	11	−0.0276	12.73	393.55	39.54
	Araneae	13	−0.0251	6.30	39.97	11.68
	Total predators	13	0.0133	4.83	34.82	9.09
	Collembola	8	−0.0297	20.01	280.27	27.81
	Myriapoda	8	−0.0021	9.05	524.01	76.54
Yellow	Cicadellidae	8	−0.0013	3.03	13.51	6.12
	Fulgoroidea	8	−0.0063	3.24	36.73	4.68
	Aphididae	8	0.0538	12.47*	68.30	9.67
	Orius spp.	8	−0.0107	7.38	35.28	10.71
	Chrysopidae	6	−0.054	28.99*	80.41	22.82
	Coccinellidae[a]	7	−0.0272	4.61	238.07	14.65
	Staphylinidae	8	−0.02	13.79	104.20	19.77
	Ichneumonidae	8	−0.0134	31.73	966.23	66.18
	Mymaridae	8	−0.0468	7.35*	30.16	4.92
	Chalcidoidea	8	−0.0173	6.25	42.42	13.46
	Chloropidae	8	−0.0202	4.60	5.65	19.78
	Muscoidea	8	0.03	13.58	107.53	24.18

[a] Only acarophagous insects are included

given trial, the means of Bt and the control treatments are statistically equal or, otherwise, that they are statistically different (the alternative hypothesis). The power under this t-test can be written as

$$Power_2 = 1 - P(T \leq t_{1-\alpha/2,2v-2}) + P(T \leq -t_{1-\alpha/2,2v-2}) \quad (6)$$

where T is a noncentral t-distributed random variable with $2v - 2$ degrees of freedom and noncentrality parameter $k = \theta\sqrt{v/2}/S_{pooled}$, $t_{1-\alpha/2,2v-2}$ is the t-Student deviate for α level of probability and $2v - 2$ degrees of freedom, v is the sample size, assuming that the Bt and the control treatments have the same sample size, S_{pooled} is the within-treatment standard deviation, pooled across the two treatment groups, and θ is the (hypothesized) *true* effect size for a given trial, respectively. Now the detectability for a given trial is

$$d_s = \frac{\theta}{m_c} 100\,\%. \quad (7)$$

On combining (6) and (7) we obtain

$$d_s = \frac{S_{pooled} k}{m_c \sqrt{v/2}} 100\,\%. \quad (8)$$

To obtain values of k, Expression (6) can only be solved numerically for a given power and a $t_{1-\alpha/2,2v-2}$ value.

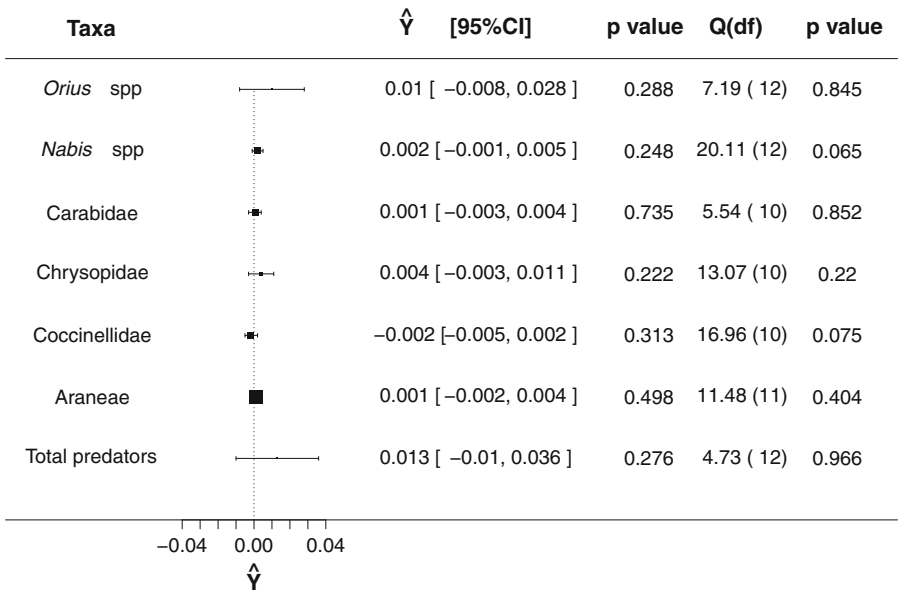

Fig. 1 Resulting effect size value $\hat{Y}$, its corresponding 95 % confidence interval [95 % CI] and p value and the value of the statistic Q, the degrees of freedom of this Q test and the corresponding p value for 7 taxa (meta-analyses) under the visual sampling technique; *square* sizes are inversely proportional to associated variance of the effect size value

Once again, on setting $\alpha = 0.05$ and a power of $0.8(\beta = 0.2)$, and solving numerically (6), we can obtain several values of k in terms of v, viz. $k = 3.0709$ and 2.3807, for $v = 3$ and 4 sample sizes, respectively: sample sizes typically used in the trials under analysis.

All the meta-analyses were computed using the *metafor* statistical package (Viechtbauer 2010) for the R statistical environment (R Development Core Team 2007).

Results

Figure 1 shows the resulting values of the common effect size $\hat{Y}$ for a given meta-analysis (i.e. taxon) together with the value of the statistic Q for visual sampling. No significant differences ($p > 0.05$) are observed in insect abundance between Bt and control (non-Bt) treatments for any taxa in visual counts. Moreover, the assumption of homogeneity is maintained for all the meta-analyses, suggesting that for a given taxon all the trials have a common effect size ($p > 0.05$) in visual counting records.

Similar results are obtained under the pitfall sampling technique (Fig. 2), in which no significant differences in insect catches ($p > 0.05$) were found between Bt and control treatments.

For the yellow trap sampling technique (see Fig. 3), no significant differences between Bt and control treatments were found for any taxon. This figure also highlights that for 3 taxa the assumption of homogeneity (i.e. a common effect size for all the studies) should be rejected ($p < 0.05$). For these 3 taxa, the effect size is not a fixed value but a random variable taking different values for each study. This finding indicates the presence of other, unincluded factors that probably explain these effect size differences between studies in yellow trap sampling. Notice that these three taxa are also the ones with most variability, which is expected because random effect meta-analyses have more variance associated with the estimated effect size. This variability consists of the within-study and between study variability, while for fixed meta-analysis only the within-study component is considered.

Table 1 shows the detectability values for the 26 meta-analyses performed [see Expressions (4) and (5)] and the maximum and minimum detectability values for all the trials considered for a given meta-analysis [see Expression (8)]. It can be seen that the meta-analysis approach dramatically improved the detectability of treatment effects compared with a single trial

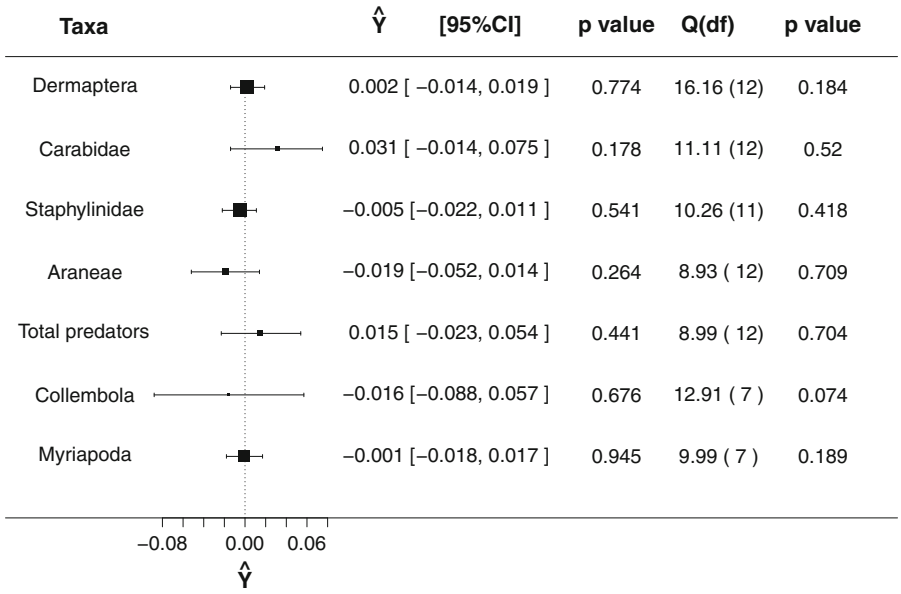

Fig. 2 As Fig. 1, but for the pitfall sampling technique

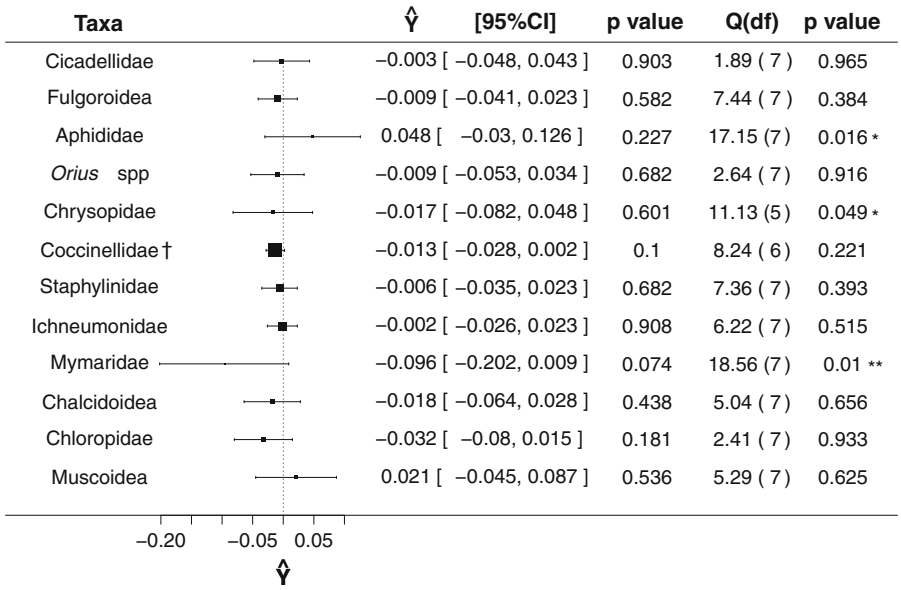

Fig. 3 As Fig. 1, but for the *yellow* sticky trap sampling technique. † Only acarophagous *insects* are included

for most taxa, regardless of the sampling technique. Notice, however, that there are three meta-analyses (i.e. taxa) in which the resulting detectability is greater than the minimum detectability obtained for a single study. Although this finding suggests that it is possible to obtain a single study with better detectability, the remaining studies show worse detectability values than those for the meta-analysis approach (compare also maximum detectability values). Moreover, these three meta-analyses have heterogeneous effect sizes. Worse detectability results are expected for the random effect meta-analyses, for which the variability associated with the meta-analysis effect size is always greater than that obtained for a fixed approach, as explained above (see for instance, Borenstein et al. 2009) [see also Expression (5)]. Table 1 also

compares meta-analysis detectability in terms of the sampling technique, highlighting that detectability depends on the taxa under analysis and does not depend greatly on the sampling technique.

Discussion

Taking all the 26 taxa and sampling techniques examined together, no significant effects of the Bt variety on NTAs were found. This finding agrees with most field studies reported in the literature, which have been reviewed by Romeis et al. (2006) and Naranjo (2009). Lack of effects in the present study is particularly important if the high detectability of the present meta-analyses is taken into account; 62 % of the taxa recorded showed detectability below 10 % and another 23 % showed detectability of between 10 and 25 %. We have published some of the results from single- or multiple-year field trials in which we report occasional effects on the predators *Orius* spp. in visual counts (de la Poza et al. 2005) and more consistent effects of Bt maize on the aphids and leafhoppers (Lumbierres et al. 2004; Pons et al. 2005). These studies were included in the present meta-analysis approaches and no-effects on the mentioned taxa were detected. This finding is especially important in the case of *Orius* spp. because the occasional effects observed by de la Poza et al. (2005) were in visual counts, the same sampling technique in which *Orius* spp. effect size was not found to be significant in its corresponding meta-analysis. In the case of differences found in aphid and leafhopper abundance in single studies, sampling was made with visual counts (Lumbierres et al. 2004; Pons et al. 2005), whereas in the studies included in the present meta-analysis the two nontarget herbivores were recorded on yellow sticky traps, which could have trapped flying individuals that were on plants other than maize; on the other hand, in sticky traps only flying adults are caught, whereas visual counts include of both nymphs and flying adults.

Global analysis of all single trials together with a meta-analysis approach dramatically improved the detection capacity of single trials for most taxa. This result is expected for a fixed effect meta-analysis in which the estimated variance of the common effect size is usually lower than that of a single trial. This is so because in a fixed effect meta-analysis each trial considered decreases the uncertainty associated with the common effect size; therefore, increasing the number of trials involves a reduction in the variance associated with the common effect size. In fact, if the associated variance for each trial were the same, the resulting variability of the common effect size would be n times smaller than that for a single study, assuming n trials for a given meta-analysis. Therefore, for 23 of 26 cases the resulting meta-analyses improve the power and therefore the detectability of the single trials. In direct contrast, for a random effect meta-analysis the resulting variance associated with the estimated global effect size does not necessarily have to be smaller than that obtained for a single trial. In this case, the global variability depends not only on the individual trial variance but also on the between-trial variability $\hat{\tau}^2$. If $\hat{\tau}^2$ is large (the trials considered in the meta-analysis are very heterogeneous), the resulting meta-analysis may not improve the results of our test compared with those obtained for a single trial. In our case, the three taxa (Mymaridae, Chrysopidae, and Aphididae on yellow sticky traps) with the highest heterogeneity (significant Q value) of effect sizes are also taxa in which the combined detectability was above the minimum detectability calculated in single trials. In particular, this global detectability was placed approximately at 30 % of all detectability for the three taxa, meaning that the global analysis gives an advantage over the separate analysis of each single trial even for the case of heterogenous trials for the majority of trials considered.

When the detectability found for the different taxa in this meta-analysis is compared with the detectability calculated in single trials (Comas et al. 2013), the benefits of the meta-analysis are also noticeable. In Comas et al. (2013), detectability was calculated on the basis of the total layout variance, that is, the variance of the residual error in the ANOVA when all treatments (not only Bt vs. non-Bt plots) were considered. The authors concluded that detectability rarely drops below 25 % in single-year trials conducted with four blocks even in the case of quite abundant taxa (Comas et al. 2013) In the present meta-analysis, only three taxa showed detectability values between 25 and 50 %, whereas for the remaining 23 taxa detectability was below 25 %. Meta-analysis requires a certain number of trials to be combined and

they are not always available; however, when available, meta-analysis may confirm results of single-year trials with a considerably greater statistical power.

In summary, the global study with a meta-analysis approach of 13 field trials carried out in Spain to measure the effects of Bt maize (with several Cry toxins) on NTAs detected no significant differences between the density or activity in Bt plots versus non-Bt plots for any of the taxa recorded and sampling techniques used. This conclusion was based on a meta-analysis of the 13 trials which would have detected relative effect sizes lower than 10 % in 62 % of the taxa recorded (16 taxa) and lower than 25 % in 85 % of the taxa recorded (22 taxa). These results are in agreement with larger field study analyses conducted in Europe with several *B. thuringiensis* Cry toxins.

Acknowledgments The results analyzed herein were obtained in field trials that were funded by the following public agencies and private companies: the Ministerio de Ciencia y Tecnología (AGF99-0782, AGL2002-00204, AGL2005-06485, AGL2011-23996), INIA and the Ministerio de Medio Ambiente, Medio Rural y Marino, and the companies Pioneer Génétique, Monsanto Agricultura España S. L. and Syngenta Seeds companies. Moreover, the editor and referees are acknowledged with thanks. Their precise comments and suggestions have clearly improved an earlier version of the manuscript.

References

Albajes R, Lumbierres B, Pons X (2009) Responsiveness of arthropod herbivores and their natural enemies to modified weed management in corn. Environ Entomol 38:944–954

Albajes R, Farinós GP, Pérez-Hedo M, de la Poza M, Lumbierres B, Ortego F, Pons X, Castañera P (2012) Post-market environmental monitoring of Bt maize in Spain: non-target effects of varieties derived from the event MON810 on predatory fauna. Spanish J Agric Res 10:977–985

Borenstein M, Hedges LV, Higgins JPT, Rothstein H (2009) Introduction to meta-analysis. Wiley, Chichester

Comas J, Lumbierres B, Pons X, Albajes R (2013) Ex ante determination of the capacity to detect treatment effects in field trials intended to assess risks of genetically modified maize on non-target arthropods. J Econ Entomol (in press)

Czarnak M, Rodríguez-Cerezo E (2010) Best practice documents for coexistence of genetically modified crops with conventional and organic farming 1. Maize crop production. EU Joint Research Centre—Institute for Prospective Technological Studies, Seville

de la Poza M, Pons X, Farinós GP, López C, Ortego F, Eizaguirre M, Castañera P, Albajes R (2005) Impact of farm-scale Bt maize on abundance of predatory arthropods in Spain. Crop Prot 24:677–684

EFSA Panel on Genetically Modified Organisms (2010) Guidance on the environmental risk assessment of genetically modified plants. EFSA J 8(11):1879

Gómez-Barbero M, Berbel J, Rodríguez-Cerezo E (2008) Bt corn in Spain—the performance of the EU's first GM crop. Nat Biotech 26:384–386

Lumbierres B, Albajes R, Pons X (2004) Transgenic Bt maize and *Rhopalosiphum padi* (Hom., Aphididae) performance. Ecol Entomol 29:309–317

Lumbierres B, Stary P, Pons X (2011) Effect of Bt maize on the plant-aphid-parasitoid tritrophic relationships. Biocontrol 56:133–143

MAGRAMA (2012) Ministerio de Agricultura, Alimentación y Medio Ambiente. Estimación de la superficie total de variedades OMG cultivadas en España en. http://www.magrama.gob.es/es/calidad-y-evaluacion-ambiental/temas/biotecnologia/2012_tcm7-220599.pdf

Marvier M, McCreedy C, Regetz J, Kareiva P (2007) A meta-analysis of effects of Bt cotton and maize on nontarget invertebrates. Science 316:1475–1477

Meissle M, Mouron P, Musa T et al (2010) Pests, pesticide use and alternative options in European maize production: current status and future prospects. J Appl Entomol 134:357–375

Naranjo SE (2009) Impacts of Bt crops on non-target invertebrates and insecticide use patterns. CAB Rev: Perspect Agric Vet Sci Nutr Nat Resour 4:1–23 http://www.cababstractsplus.org/cabreviews

Perry JN, Rothery P, Clark SJ, Heard MS, Hawes C (2003) Design, analysis and statistical power of the farm-scale evaluations of genetically modified herbicide-tolerant crops. J Appl Ecol 40:17–31

Pons X, Lumbierres B, López C, Albajes R (2005) Abundance of non-target pests in transgenic Bt-maize: a farm scale study. Eur J Entomol 102:73–79

Romeis J, Meissle M, Bigler F (2006) Transgenic crops expressing *Bacillus thuringiensis* toxins and biological control. Nat Biotechnol 24:63–71

Romeis J, Bartch D, Bigler F et al (2008) Assessment of risk of insect-resistant transgenic crops to nontarget arthropods. Nat Biotechnol 26:203–208

R Development core team (2007) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org

Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. J Stat Softw 36:1–48

Whitehead A (2002) Meta-analysis of controlled clinical trials. Wiley, Chichester

Wolfenbarger LL, Naranjo SE, Lundgren JG, Bitzer RJ, Watrud LS (2008) Bt crops effects on functional guilds of non-target arthropods: a meta-analysis. PLoS ONE 3(5):e2118