Investigation of growth regulation by maize benzoxazinoid breakdown products
Jander, G. Boyce Thompson Institute
Chamovitz, A.D. Tel Aviv University

Project award year: 2015
Three year research project
Investigation of growth regulation by maize benzoxazinoid breakdown products

Introduction

Previous research had suggested that benzoxazinoids, a class of defensive metabolites found in maize, wheat, rye, and wild barley, are not only direct insect deterrents, but also influence other areas of plant metabolism. In particular, the benzoxazinoid 2,4-dihydroxy-7-methoxy-2H-1,4-benzoxa-zin-3(4H)-one (DIMBOA) was implicated in: (i) altering plant growth by interfering with auxin signaling, and (ii) leading to the induction of gene expression changes and secondary plant defense responses. The overall goal of this proposal was to identify mechanisms by which benzoxazinoids influence other aspects of plant growth and defense. Specifically, the following hypotheses were proposed to be tested as part of an approved BARD proposal:

1. Benzoxazinoid breakdown products directly interfere with auxin perception
2. Global changes in maize and barley gene expression are induced by benzoxazinoid activation.
3. There is natural variation in the maize photomorphogenic response to benzoxazinoids.

Although the initial proposal included experiments with both maize and barley, there were some technical difficulties with the proposed transgenic barley experiments and most of the experimental results were generated with maize.

Summary of major findings

Previous research by other labs, involving both maize and other plant species, had suggested that DIMBOA alters plant growth by interfering with auxin signaling. However, experiments conducted in both the Chamovitz and the Jander labs using Arabidopsis and maize, respectively, were unable to confirm previously published reports of exogenously added DIMBOA effects on auxin signaling. Nevertheless, analysis of bx1 and bx2 maize mutant lines, which have almost no detectable benzoxazinoids, showed altered responses to blue light signaling. Transcriptomic analysis of maize mutant lines, variation in inbred lines, and responses to exogenously added DIMBOA showed alteration in the transcription of a blue light receptor, which is required for plant growth responses. This finding provides a novel mechanistic explanation of the trade-off between growth and defense that is often observed in plants.

Experiments by the Jander lab and others had demonstrated that DIMBOA not only has direct toxicity against insect pests and microbial pathogens, but also induces the formation of callose in both maize and wheat. In the current project, non-targeted metabolomic assays of wildtype maize and mutants with defects in benzoxazinoid biosynthesis were used to identify unrelated metabolites that are regulated in a benzoxazinoid-dependent manner. Further investigation identified a subset of these DIMBOA-responsive compounds as catechol, as well as its glycosylated and acetylated derivatives.

Analysis of co-expression data identified indole-3-glycerol phosphate synthase (IGPS) as a possible regulator of benzoxazinoid biosynthesis in maize. In the current project, enzymatic activity of three predicted maize IGPS genes was confirmed by heterologous expression. Transposon knockout mutations confirmed the function of the maize genes in benzoxazinoid biosynthesis. Sub-cellular localization studies showed that the three maize IGPS proteins are co-localized in the plastids, together with BX1 and BX2, two previously known enzymes of the benzoxazinoid biosynthesis pathway.

Implications

Benzoxazinoids are among the most abundant and effective defensive metabolites in maize, wheat, and rye. Although there is considerable within-species variation in benzoxazinoid content, very little is known about the regulation of this variation and the specific effects on plant growth and defense. The results of this research provide further insight into the complex functions of maize benzoxazinoids, which are not only toxic to pests and pathogens, but also regulate plant growth and other defense responses. Knowledge gained through the current project will make it possible to engineer benzoxazinoid biosynthesis in a more targeted manner to produce pest-tolerant crops without negative effects on growth and yield.
Summary Sheet

Publication Summary

<table>
<thead>
<tr>
<th>PubType</th>
<th>IS only</th>
<th>Joint</th>
<th>US only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review Article</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Reviewed</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Training Summary

<table>
<thead>
<tr>
<th>Trainee Type</th>
<th>Last Name</th>
<th>First Name</th>
<th>Institution</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postdoctoral Fellow</td>
<td>Richter</td>
<td>Annett</td>
<td>Boyce Thompson Institute</td>
<td>USA</td>
</tr>
<tr>
<td>Ph.D. Student</td>
<td>Katz</td>
<td>Ella</td>
<td>Tel Aviv University</td>
<td>Israel</td>
</tr>
<tr>
<td>Ph.D. Student</td>
<td>Berliner</td>
<td>Mark</td>
<td>Tel Aviv University</td>
<td>Israel</td>
</tr>
</tbody>
</table>
Contribution of the collaboration

The Jander and Chamovitz labs were actively involved in this collaborative research project.

1. The two labs worked collaboratively to investigate DIMBOA interactions with auxin signaling in Arabidopsis, maize, and barley. Although we found that we could not confirm previously published reports of this phenomenon, this research led to the identification of DIMBOA-dependent transcriptional changes that alter maize photomorphogenesis.

2. With advice from Jander, graduate student Mark Berliner in the Chamovitz lab worked to transform barley with the maize Bx6 and Bx7 genes to make its defenses more like those of maize and wheat. Unfortunately, technical difficulties precluded this project from coming to a conclusion.

3. Annett Richter, a postdoc in the Jander lab investigated changes in the induced defense responses of maize lines that are defective in benzoxazinoid biosynthesis. Analysis of these plant responses occurred in consultation with the Chamovitz lab.

4. Chamovitz visited the Jander lab at the Boyce Thompson Institute in the summer of 2017 for research discussions.

5. Jander visited the Chamovitz lab at Tel Aviv University in the spring of 2019 for research discussions.
Investigation of growth regulation by maize benzoxazinoid breakdown products

BARD Award US-4846-15C

Georg Jander, Boyce Thompson Institute

Danny Chamovitz, Tel Aviv University

Achievements

The goal of this research was to investigate the extended function of benzoxazinoids, an important class of defensive metabolites found in maize, wheat, rye, and wild barley. Specifically, we proposed to investigate: (i) the role benzoxazinoids in auxin signaling, and (ii) the role of benzoxazinoids in regulating the activation of other maize defense responses. BARD-funded research resulted in novel findings in three main research areas:

1. Investigation of photomorphogenic responses to benzoxazinoids

Several older publications have suggested a direct effect of benzoxazinoids, in particular 2,4-dihydroxy-7-methoxy-2H-1,4-benzoazin-3(4H)-one (DIMBOA) and its breakdown products, on auxin signaling. However, research in the Jander and Chamovitz labs using more modern molecular methods and mutant lines failed to confirm this observation using either maize or Arabidopsis. Nevertheless, we observed photomorphogenic changes in maize that were associated with benzoxazinoid production. These were ascribed to differential transcriptional regulation of a maize *NPH3* gene that is required for blue light perception and the resulting growth changes. This result still needs to be confirmed with mutant maize lines, specifically to determine whether a knockout mutation in the *NPH3* gene prevents DIMBOA-regulated maize growth responses.

Our results are significant because they provide a previously unknown link between maize defense responses and plant growth. Although trade-offs between plant growth and defense are commonly reported, the mechanisms by which this occurs are poorly understood. A more thorough investigation of this process is agriculturally relevant, because manipulation of the growth/defense tradeoff may lead to strategies that increase endogenous plant defenses without negative influences on plant growth.

2. Effects of benzoxazinoids on the induction of other plant defense responses
Previous work by the Jander lab and others had demonstrated that DIMBOA induces the formation of callose as a defense response in both maize and wheat. We hypothesized, that in addition to callose induction, benzoxazinoids would induce other defense-related responses in maize. With a combination of transcript profiling and metabolite profiling, we were able to show that there are broad changes in maize metabolism in the absence of benzoxazinoids, and that some of the observed changes could be induced in a benzoxazinoid-dependent manner. Furthermore, we went on to demonstrate that catechol, together with its glycosylated and acetylated derivatives was induced in a benzoxazinoid-dependent manner.

To date a clear defense signaling function of an otherwise directly defense-related metabolites has only been demonstrated for benzoxazinoids and glucosinolates, a crucifer-specific class of defensive metabolites. Therefore, the effect that we are investigating is scientifically interesting and worthy of investigation. It is likely also some of the first evidence of a broader phenomenon of incipient signaling pathways that is likely to be common in plant biology. The release of mobile defense-related metabolites is a clear sign of attack, and it is quite reasonable to assume that plants can use this as an endogenous signal to activate other defense responses.

Targeted regulation of plant defense, i.e. only turning on the right defenses at the right time, is a goal of digital agriculture. Results from the current research provide new insight into the regulation of maize defenses, as well as leading to the discovery of previously unknown maize defensive metabolites. These discoveries may make it possible to breed maize with a higher level of natural pest tolerance.

3. Indole-3-glycerol phosphate synthase as a branchpoint between primary and secondary metabolism

The canonical pathway of maize benzoxazinoid biosynthesis begins with BX1, a maize homolog of the *E. coli* tryptophan synthase alpha subunit, which converts indole-3-glycerolphosphate to indole. However, our analysis of maize gene co-expression indicated that the branchpoint between primary and secondary metabolism in maize is actually at the level of indole-3-glycerolphosphate (IGPS). Unlike Arabidopsis and most other dicots, maize has three *IGPS* genes. As part of the current project, we investigated the maize IGPS *in vitro* enzymatic activity, sub-cellular localization, gene expression regulation, and *in vivo* effects of knockout mutations.
This demonstrated that the three IGPS enzymes have distinct roles in maize primary metabolism (tryptophan biosynthesis) and secondary metabolism (release of free indole and benzoxazinoid biosynthesis). The results of this research are described in a manuscript that is in preparation.
Changes to the Original Research Plan

The original research plan included experiments to make genetically engineered barley with a benzoxazinoid content more like that of maize and wheat. The purpose was to determine whether the secondarily evolved benzoxazinoids from maize and wheat provide enhanced defense. However, technical difficulties with cloning of benzoxazinoid genes and barley transformation precluded successful completion of these experiments. Instead, the focus of our work was placed primarily on maize research.
<table>
<thead>
<tr>
<th>Status</th>
<th>Type</th>
<th>Authors</th>
<th>Title</th>
<th>Journal</th>
<th>Vol:pg</th>
<th>Year</th>
<th>Cou n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published</td>
<td>Review Article</td>
<td>Zhou S, Richter A, and Jander G</td>
<td>Beyond defense: Multiple function of benzoxazinoids in maize metabolism</td>
<td>Plant &amp; Cell Physiology</td>
<td>59:1528–1537</td>
<td>2018</td>
<td>US only</td>
</tr>
</tbody>
</table>
Investigation of growth regulation by maize benzoxazinoid breakdown products

BARD Award US-4846-15C
Georg Jander, Boyce Thompson Institute
Danny Chamovitz, Tel Aviv University

Appendix

<table>
<thead>
<tr>
<th>Page</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Table of contents</td>
</tr>
<tr>
<td>2-11</td>
<td>Publication, Zhou et al, 2018</td>
</tr>
<tr>
<td>12-23</td>
<td>Publication, Springer et al, 2018</td>
</tr>
<tr>
<td>24-47</td>
<td>Manuscript in preparation: IGPS activity in maize</td>
</tr>
<tr>
<td>48-49</td>
<td>Notable results: regulation of photomorphogenesis by benzoxazinoids</td>
</tr>
<tr>
<td>50-51</td>
<td>Notable results: regulation of secondary metabolism by benzoxazinoids</td>
</tr>
</tbody>
</table>
Beyond Defense: Multiple Functions of Benzoxazinoids in Maize Metabolism

Shaoqun Zhou¹,², Annett Richter¹ and Georg Jander¹,*

¹Boyce Thompson Institute for Plant Research, 533 Tower Road, Ithaca, NY 14853, USA
²Plant Biology Section, School of Integrated Plant Science, Cornell University, Ithaca, NY 14853

*Corresponding author: E-mail, gj32@cornell.edu; Fax, (1) 607-254-1242.

(Received January 6, 2018; Revised March 18, 2018)

Benzoxazinoids are a class of indole-derived plant metabolites that function in defense against numerous pests and pathogens. Due to their abundance in maize (Zea mays) and other important cereal crops, benzoxazinoids have been the subject of extensive research for >50 years. Whereas benzoxazinoids can account for 1% or more of the dry weight in young seedlings constitutively, their accumulation in older plants is induced locally by pest and pathogen attack. Although the biosynthetic pathways for most maize benzoxazinoids have been identified, unanswered questions remain about the developmental and defense-induced regulation of benzoxazinoid metabolism. Recent research shows that, in addition to their central role in the maize chemical defense repertoire, benzoxazinoids may have important functions in regulating other defense responses, flowering time, auxin metabolism, iron uptake and perhaps aluminum tolerance. Investigation of natural variation in maize benzoxazinoid accumulation, which is greatly facilitated by recent genomics advances, will have a major impact in this research area by leading to the discovery of previously unknown genes and functions of benzoxazinoid metabolism.

Keywords: Benzoxazinoid • Defense • DIMBOA • Maize • Signaling.

Abbreviations: BOA, benzoxazolin-2-one; DIBOA, 2,4-dihydroxy-1,4-benzoxazin-3-one; DBMBOA, 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one; DMBOA, 6,7-dimethoxy-2-benzoxazolinone; GWAS, genome-wide association study; HDMBOA, 2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one; HDM2BOA, 2-dihydroxy-4,7,8-trimethoxy-1,4-benzoxazin-3-one; HMBOA, 2-hydroxy-7-methoxy-1,4-benzoxazin-3(4H)-one; IGL, indole-3-glycerolphosphate lyase; IGPS, indole-3-glycerolphosphate synthase; MBOA, 6-methoxy-benzoxazolin-2-one; NAM, nested association mapping; ODD, oxoglutarate-dependent dioxygenase; OMT, O-methyltransferase; QTL, quantitative trait locus; TRIBOA, 2,4,7-trihydroxy-2H-1,4-benzoxazin-3(4H)-one; TRIMBOA, 2,4,7-trihydroxy-8-methoxy-1,4-benzoxazin-3-one; TSA, tryptophan synthase alpha subunit.

Introduction

Benzoxazinoids, which are found in maize (Zea mays), wheat (Triticum spp.), rye (Secale cereale) and other wild and cultivated Poaceae (Kokubo et al. 2017), are among the most agriculturally relevant groups of plant specialized metabolites. This family of approximately 20 compounds shares a 2-hydroxy-2H-1,4-benzoxazin-3(4H)-one skeleton (Fig. 1; HBOA). In plants such as maize, the predominant benzoxazinoids can account for 0.1–0.3% of fresh weight in some tissues, particularly in young seedlings (Dafao et al. 2011, Glaser et al. 2011, Meihls et al. 2013). Since their discovery in the 1950s, benzoxazinoids have been associated with biochemical defense against a variety of biotic stresses, including insect herbivores, microbial pathogens and competing plant species (reviewed in Niemeyer 2009). These interesting ecological functions of benzoxazinoids, which have shed light on their potential for crop protection and pest management, have attracted extensive research interest. Due to the well-developed genetic resources that are available for maize, many of the recent advances in benzoxazinoid research have been made with this species. Here we aim to provide an updated review of benzoxazinoid biosynthesis and function, with an emphasis on describing new developments from the past 3 years.

Biosynthesis of Benzoxazinoids

Efforts to elucidate the biosynthetic pathway of benzoxazinoids date back to the 1960s. Almost 40 years of active research led to the identification of the entire pathway leading from indole to DIBOA (2,4-dihydroxy-1,4-benzoxazin-3-one), which is primarily a metabolic intermediate in maize benzoxazinoid biosynthesis, but the main defensive end-product in rye and wild barley (Frey et al. 1997; Fig. 1). It is likely that the DIBOA biosynthetic pathway represents the first completely characterized pathway for the biosynthesis of any compound in plant secondary metabolism.

The first committed step of benzoxazinoid biosynthesis is generally considered to be the production of indole from indole-3-glycerolphosphate by BX1 in the chloroplast. BX1 has the same enzymatic activity as the tryptophan synthase α-subunit, but the indole products of the two enzymes are used in different ways by maize plants. Although bx1 knockout mutations result in a very low level of benzoxazinoids (Frey et al. 1997, Tzin et al. 2017), the plants appear morphologically normal, indicating that tryptophan biosynthesis is largely unaffected. Two other maize proteins, IGL1 (indole-3-
glycerolphosphate lyase 1) and TSA1 (tryptophan synthase alpha-subunit 1; Kriechbaumer et al. 2008), synthesize indole and probably compensate for the lack of activity in the bx1 mutant line. Recent clustering analysis of maize gene expression data suggests that, whereas BX1 has a primary function in benzoxazinoid biosynthesis, TSA1 and IGL1 contribute to tryptophan and free indole biosynthesis, respectively (Wisecaver et al. 2017). The pathway-specific synthesis of indole may be due to a different subcellular localization of the three enzymes, or perhaps channeling of indole directly to the tryptophan synthase for tryptophan biosynthesis. Additionally, temporal and spatial differences in the expression of these genes may also contribute to the pathway-specific indole synthesis. The residual accumulation of DIMBOA-Glc (2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside) and HDMBOA-Glc (2-hydroxy-4,7-dimethoxy-1,4-benzoxazin-3-one glucoside) in the W22 bx1 mutant (Tzin et al. 2017) may indicate some indole leakage toward benzoxazinoid biosynthesis from the other two pathways.

Clustering analysis of maize gene expression also showed that one of three predicted indole-3-glycerolphosphate synthases (IGPS1; GRMZM2G106950; Fig. 1) is co-expressed with the known benzoxazinoid biosynthetic genes (Wisecaver et al. 2017). IGPS1 is also the only one of the three maize IGPS genes that is strongly induced by Spodoptera exigua (beet army-worm) feeding (Tzin et al. 2017). Thus, it is possible that metabolic channeling toward benzoxazinoid biosynthesis occurs one step earlier than BX1.

The indole produced by BX1 is oxidized at four positions by a series of Cyt P450-dependent monooxygenases (BX2–BX5, all of which are in the CYP71C subfamily) in the endoplasmic reticulum, producing the simplest functional benzoxazinoid, DIBOA (Frey et al. 1997). Although each of these oxygenation reactions is catalyzed by a separate Cyt P450, there may be some overlap in substrate specificity. In particular, a bx2 transposon knockout mutant still produces residual amounts of DIMBOA-Glc and HDMBOA-Glc (Tzin et al. 2017), suggesting that BX3, BX4 or BX5 is able to catalyze the BX2 reaction at a lower level. Double mutants of these genes could be used to test the hypothesis of enzymatic redundancy. However, it is also possible that an as yet uncharacterized maize CYP71C protein can have overlapping enzymatic functions with BX2.

Fig. 1 Major maize benzoxazinoids and enzymes catalyzing their biosynthesis.
Two partially redundant UDP-glucosyltransferases, BX8 and BX9, attach a glucose moiety to the cytotoxic DIBOA and DIMBOA to produce the less toxic DIBOA-Glc and DIMBOA-Glc, respectively (von Rad et al. 2001). Co-expression analysis shows that another predicted maize glucosyltransferase (GRMZM2G085854) is in a transcriptional module associated with benzoxazinoid biosynthesis (expressed together with BX1–BX5; Wisecaver et al. 2017).

As this gene is also physically located within the cluster of benzoxazinoid biosynthetic genes on maize chromosome 4, it might also contribute to benzoxazinoid biosynthesis. It is possible that the different benzoxazinoid glucosyltransferases have different functions in either initial synthesis of benzoxazinoids or the detoxification of breakdown products after activation by glucosidases.

DIBOA-Glc is further oxidized by a 2-oxoglutarate-dependent dioxygenase (2-ODD, BX6) and methylated by an O-methyltransferase (OMT, BX7) to produce DIMBOA-Glc (Jonczyk et al. 2008). Whereas the originally identified BX6 mutant line has almost no DIMBOA-Glc (Jonczyk et al. 2008), a bx6 transposon knockout mutant in the W22 genetic background still accumulates about 40% of the DIMBOA-Glc of wild-type W22 (Tzin et al. 2015a). This suggests that, at least in the W22 genetic background, some other maize enzyme also has this 2-ODD activity.

More recently, quantitative genetics approaches have been used to characterize the later steps in maize benzoxazinoid biosynthesis. Genetic mapping of HDMBOA-Glc abundance using recombinant inbred lines of the maize nested association mapping (NAM) population (McMullen et al. 2009) identified a cluster of three genes (BX10, BX11 and BX12) that encode OMTs catalyzing the conversion of DIMBOA-Glc to HDMBOA-Glc (Meihls et al. 2013). Creation of near-isogenic lines demonstrated that a CACTA-family transposon insertion in the BX13 gene (Handrick et al. 2016). This variation in benzoxazinoid content is also reflected in the wide variation in the expression of BX genes across the different inbred lines of the NAM population (http://qteller.com/ (April 7, 2018, date last accessed)). Quantitative trait mapping approaches have provided some insight into the underlying genetic basis of natural variation in benzoxazinoid content. Low constitutive levels of HDMBOA-Glc in many temperate maize inbred lines are caused by a transposon insertion in the BX12 gene (Meihls et al. 2013), whereas the absence of DIM2BOA-Glc and HDM2BOA-Glc in inbred line Oh43 is caused by a start codon deletion in the BX13 gene (Handrick et al. 2016).

Benzoxazinoid glucosides are stored in vacuoles of maize cells. Upon tissue disruption, these glucosides are exposed to separately localized glucosidases (Esen 1992, Czajek et al. 2001, Nikus et al. 2001), which cleave off the glucosyl groups, producing the bioactive aglucones and further breakdown products. The process of benzoxazinoid activation and the chemical functions of breakdown products have been described extensively in two recent reviews (Wouters et al. 2016a, Wouters et al. 2016b) and will not be covered here.

**Genetic and Developmental Variation in Constitutive Benzoxazinoid Content**

Assays of the 26 genetically diverse parental lines of the maize NAM population have demonstrated considerable constitutive variation in both leaf and root benzoxazinoid content, and very little correlation in benzoxazinoid profiles between the two tissue types in the same inbred line (Meihls et al. 2013, Handrick et al. 2016). This variation in benzoxazinoid content is also reflected in the wide variation in the expression of BX genes across the different inbred lines of the NAM population (http://qteller.com/ (April 7, 2018, date last accessed)).
distal *cis*-element (DICE), that is 140 kbp upstream of *BX1*, regulates expression of this gene and thereby the rate at which benzoxazinoid content drops during the development of maize seedlings (Zheng et al. 2015). Interestingly, the DICE region regulates only *BX1* expression but not that of either the intervening *BX8* gene or the *BX2* gene, which is only 2.5 kbp downstream of *BX1* in the inbred line B73 genome (Schnable et al. 2009). Differences in the constitutive benzoxazinoid content between B73 and Mo17 seedlings have also been associated with four transcription factors, which have expression that is correlated with the *BX1* expression pattern in the B73 and Mo17 inbred lines (Song et al. 2017).

These recently identified genes all represent ‘low-hanging fruit’ in the identification of natural variation that influences benzoxazinoid biosynthesis in maize. In the case of *BX12* and *BX13*, gene identification was greatly facilitated by natural knockout mutations that segregate as Mendelian loci. There are undoubtedly many more examples of *cis*- and *trans*-regulation of gene expression that influence constitutive and developmental variation in benzoxazinoid abundance. With more sequenced maize genomes and genome-wide association studies (GWASs) using large numbers of inbred lines (Fint-Garcia et al. 2005, Samayoa et al. 2015), it is likely that other regulatory loci will be identified in rapid succession.

Variation in benzoxazinoid content in different tissue types could also be explained by long-distance transport of benzoxazinoids after they are produced. The presence of benzoxazinoids in the leaf apoplast and phloem exudates (Ahmad et al. 2011) suggests transport between different plant tissues. As described above, two benzoxazinoid-specific glucosyltransferases, *BX8* and *BX9*, are thought to be localized to the cytosol, acting on newly synthesized DIBOA molecules released from the endoplasmic reticulum. This glycosylation reaction is important to stabilize the benzoxazinoid molecule to avoid autoxidation to plant cells. *BX6*, which encodes a 2-ODD that is required to synthesize DIMOBA-Glc (Fig. 1), takes only DIBOA-Glc but not DIBOA aglucone as a substrate. Therefore, it has been hypothesized that benzoxazinoid molecules are released into the apoplasm as glucosides, and later activated by extracellular glucosidases (Ahmad et al. 2011). However, further study is required to test this hypothesis and elucidate the process of benzoxazinoid export from cells into the apoplastic space, as root exudates (Neal et al. 2012, Neal and Ton 2013) or as components of the waxy surface layer of leaves (Hedin et al. 1993).

A corollary to the wide variation in maize benzoxazinoid content is that it is not possible to make broad generalizations based on analyzing only one maize inbred or hybrid line. Descriptions of previous findings such as DIMOBA-Glc being the predominant benzoxazinoid in leaves (Cambier et al. 2000) and HDMBOA-Glc being predominant in roots (Robert et al. 2012) should always be qualified with a statement such as ‘in inbred line X’. There are certainly maize lines where DIMOBA-Glc is not predominant in leaves and HDMBOA-Glc is not predominant in the roots (Meihls et al. 2013, Hendrick et al. 2016). Similarly, the phenotypic effect of *bx6* knockout mutations depends on the maize genetic background (Jonczyk et al. 2008, Tzin et al. 2015a). Another possible source of variation in benzoxazinoid data interpretation among researchers comes from differences in the sample preparation methods, which can vary in their extraction efficiency, the extent of benzoxazinoid degradation and the chromatographic separation of different benzoxazinoid types (Pedersen et al. 2017a, Pedersen et al. 2017b).

### Environmentally Induced Changes in Benzoxazinoid Content

As a class of defense-related specialized metabolites, benzoxazinoids are often considered to be phytoanticipins, compounds that are constitutively produced by plants even in the absence of an imminent threat. However, constitutively high levels of benzoxazinoids are generally only found in maize seedlings, and even in seedlings it has been shown that insect feeding can induce further benzoxazinoid accumulation and modification (Köhler et al. 2015, Tzin et al. 2017). Leaves of mature maize plants have relatively low benzoxazinoid levels, but these are subject to rapid increases in response to attack by both pests and pathogens (Huffaker et al. 2011, Maag et al. 2016).

Transcriptomic and metabolomic studies show induced changes in benzoxazinoid metabolism in response to insect herbivory (Tzin et al. 2015a, Tzin et al. 2017). Younger maize leaves show stronger benzoxazinoid induction after *Spodoptera littoralis* (African cotton leafworm) feeding than older leaves on the same plant (Köhler et al. 2015). Induced benzoxazinoid accumulation is highly localized to the site of caterpillar feeding and persists for several days (Maag et al. 2016). Both *S. littoralis* and *Ostrinia furnicularis* (Asian corn borer) grew more slowly on maize with prior insect damage (Maag et al. 2016, Guo et al. 2017). The resistance effect was eliminated in a *bx1* mutant line, indicating that induced maize defenses against these caterpillars require benzoxazinoid synthesis.

Infection with fungal pathogens (Oikawa et al. 2004, Huffaker et al. 2011), treatment with jasmonic acid (Oikawa et al. 2001, Oikawa et al. 2002) and feeding by several lepidopteran species, including *O. furnicularis* (Guo et al. 2017), *Ostrinia nubilalis* (European corn borer; Dafae et al. 2011, Dafae et al. 2013), *Spodoptera frugiperda* (fall armyworm; Gläuser et al. 2011), *S. exigua* (Tzin et al. 2017) and *S. littoralis* (Gläuser et al. 2011, Köhler et al. 2015, Maag et al. 2016), induces the methylation of DIMOBA-Glc to form HDMBOA-Glc. In the case of *S. exigua* feeding, expression of the *BX10* and *BX11* OMTs is induced within 1 h of caterpillar addition to the plants, followed by increased expression of earlier genes in the benzoxazinoid pathway after 4 h (Tzin et al. 2017). This inducible conversion may enhance maize resistance to insect herbivores, because the HDMBOA aglucone is degraded to MBOA more rapidly than the DIMOBA aglucone (Oikawa et al. 2004, Gläuser et al. 2011, Meihls et al. 2013). Additionally, re-glycosylation of activated benzoxazinoids as a detoxification mechanism in *S. frugiperda* is less effective in counteracting HDMBOA-Glc rather than DIMOBA-Glc activation by glucosidases (Gläuser et al. 2011).
Given the presence of benzoxazinoids in phloem exudates (Ahmad et al. 2011), it is possible that they are not only produced locally, but are also transported to the site of insect feeding. Benzoxazinoid transport is also suggested by the low level of correlation between transcript levels of the BX10, BX11 and BX14 methyltransferase genes and the relative abundance of HDMBOA-Glc and HDM2BOA-Glc in S. littoralis-infested B73 leaf tissue (Maag et al. 2016). In particular, the concentration of these benzoxazinoids in leaf tips was lower than what would be expected based on the transcript levels. However, this metabolite distribution could also be explained by post-transcriptional regulation of enzyme activity.

The soil bacterium Pseudomonas putida, which is generally considered beneficial to plants, is attracted to DIMBOA in the rhizosphere (Neal et al. 2012). One beneficial effect of the interaction with maize is defense priming, whereby jasmonate-induced defenses in leaves are turned on more rapidly if the roots are colonized by P. putida. Although this effect is absent in igl1 bx1 mutant plants that lack benzoxazinoids (Neal and Ton 2013), it is not known whether foliar defense priming involves transport of benzoxazinoids from the roots to the leaves, different recruitment of P. putida in benzoxazinoid-deficient plants or simply a different defense physiology of leaves in the absence of benzoxazinoids.

In addition to the biotic stresses described above, benzoxazinoid accumulation can also be regulated by diverse abiotic stresses, including temperature, drought, nitrogen deficiency, photoperiod, light intensity and UV radiation (reviewed in Niemeyer 2009). However, the molecular mechanisms underlying these physiological observations are yet to be elucidated.

**Benzoxazinoids as Defense Signaling Molecules**

Like indole-3-ylmethylglucosinolate, which has been shown to induce callose formation as a pathogen defense response in Arabidopsis (Clay et al. 2009), benzoxazinoids induce the formation of callose in maize during pathogen infection and insect feeding. A bx1 igl1 double mutant, in comparison with an igl1 single mutant, was used to demonstrate the influence of benzoxazinoids in callose formation induced by chitosan, a fungal elicitor (Ahmad et al. 2011). Notably, exogenous addition of DIMBOA, but not HDMBOA-Glc, triggered the formation of callose in maize leaves (Ahmad et al. 2011).

Different alleles of BX12, which converts DIMBOA-Glc to HDMBOA-Glc, were identified as a cause of natural variation in Rhopalosiphipum maidis (corn leaf aphid) resistance (Meihls et al. 2013). Even though HDMBOA-Glc is more toxic to R. maidis in vitro, these aphids grew less well on plants with higher DIMBOA-Glc levels. Induction of callose by DIMBOA could enhance aphid resistance by blocking phloem sieve elements and/or by making cell walls more difficult for aphid styllets to penetrate. Variation in the expression of BX1, a genetic basis for differences in the DIMBOA-Glc content of B73 and Mo17 seedlings (Zheng et al. 2015), also affects R. maidis resistance (Betsiashvili et al. 2015). Similar differences in callose-dependent aphid resistance on the B73 and Mo17 inbred lines were observed with a different aphid species, Rhopalosiphipum padi (bird cherry-oat aphid; Song et al. 2017).

The predicted functions of HDMBOA-Glc in preventing caterpillar feeding and DIMBOA-Glc in triggering callose formation as an anti-aphid response lead to potential trade-offs in maize defense. In particular, caterpillar-induced conversion of DIMBOA-Glc to HDMBOA-Glc was predicted to cause decreased aphid resistance. Support for this hypothesis comes from experiments in which S. exigua-induced resistance to R. maidis was mapped as a quantitative trait using B73×Ky21 recombinant inbred lines (Tzin et al. 2015b). One of the identified quantitative trait loci (QTLs) coincides with the previously identified BX10–12 region on chromosome 1 (Meihls et al. 2013), suggesting that caterpillar-induced methylation of DIMBOA-Glc to form HDMBOA-Glc leads to decreased aphid resistance.

Although the induction of callose by DIMBOA in maize is by now well established, the mechanism by which this occurs is not known. DIMBOA could induce callose synthases at the transcriptional level, or perhaps activate pre-existing callose synthases as a more rapid response to aphid feeding. DIMBOA, a small phloem-mobile molecule that becomes more abundant in response to insect feeding, could be an effective signaling molecule that provides evidence of herbivore attack and induces secondary plant responses. However, it is not yet known whether other plant defenses, in addition to callose formation, are triggered by the release of DIMBOA due to insect herbivory or pathogen infection.

**The Sweet Taste of Benzoxazinoids**

Specialized insect herbivores are often attracted by defensive metabolites that are characteristic of their preferred host plants. Among the best-studied examples of this phenomenon is the attraction of crucifer-feeding specialists such as Pieris rapae (cabbage butterfly) and Plutella xylostella (diamondback moth) to glucosinolates for feeding and oviposition (Renwick et al. 1992, Hughes et al. 1997, De Vos et al. 2008). Similarly, Diabrotica virgifera (corn rootworm) larvae, which specialize in feeding on maize roots, use benzoxazinoids as chemical cues to locate more nutritious crown roots as feeding sites (Robert et al. 2012).

Some specialist insect herbivores co-opt host plant defensive chemicals for their own defense against predators. For instance, Brevicoryne brassicae (cabbage aphids) and Danausplexippus (monarch butterflies), respectively, sequester glucosinolates from crucifers and cardiac glycosides from milkweed (Jones et al. 2001, Dobler et al. 2012). A similar defensive strategy was recently identified in D. virgifera, which sequesters HDMBOA-Glc and activates this metabolite to form MBOA as a defense against infection by entomopathogenic nematodes and their symbiotic bacteria (Robert et al. 2017). MBOA that is formed by spontaneous degradation of DIMBOA during root feeding is glycosylated by D. virgifera to form MBOA-Glc, and nematodes are repelled by MBOA-Glc that is excreted in the larval frass (Robert et al. 2017).
It is possible that benzoxazinoids, in particular DIMBOA, also function as feeding stimulants for non-specialized herbivores. Maize benzoxazinoids were originally identified in 1962 as the ‘corn sweet substance’ (Hamilton et al. 1962). For human taste buds, DIMBOA is about 400 times sweeter than sucrose, and Hamilton et al. (1962) reported that: ‘Etiolated corn seedlings were noted to have a very sweet saccharin-like taste’. Since other mammals are also likely to find DIMBOA sweet tasting, it is possible that higher DIMBOA content, despite providing increased toxicity, might make maize seedlings more attractive for mammalian herbivores. Many insects, for instance bees and aphids, are attracted by sugars in their diet. However, it is not known whether DIMBOA also has a sugar-like effect on insect taste receptors and thereby could serve as a feeding stimulant for maize-eating insects.

Possible Regulation of Flowering Time by Benzoxazinoids

Analysis of 279 maize inbred lines from the Goodman Diversity Panel (Flint-Garcia et al. 2005) confirmed the previously identified pattern (Meihls et al. 2013) that tropical maize varieties have higher seedling HDMBOA-Glc content than temperate maize varieties (Fig. 2). Furthermore, analysis of a large maize×teosinte recombinant inbred line population showed that the Bx12 locus was under strong selection as maize cultivation spread into more temperate environments (Wang et al. 2018). None out of 25 tested teosinte isolates had the CACTA transposon insertion in Bx12, which is found in most temperate maize varieties, suggesting that this knockout mutation was selected during domestication (Wang et al. 2018).

It is tempting to speculate that selection for different types of herbivore resistance accounts for the observed difference in benzoxazinoid content between tropical and temperate maize varieties (Fig. 2; Supplementary Table S1). For instance, tropical maize seedlings with high HDMBOA-Glc content could be selected to have greater resistance to caterpillars, which can devastate newly planted maize in tropical areas of Mexico. Conversely, temperate maize with higher DIMBOA-Glc content has higher aphid resistance (Meihls et al. 2013). Aphids tend to be a greater problem in temperate areas and do not survive well at high temperatures in the tropics.

Earlier flowering time and differing day-length sensitivity are among the strongest selective pressures that were applied to maize as cultivation moved to more temperate climates (Swarts et al. 2017). In a recent genetic analysis of 4,471 maize landraces, BX12 was identified as a locus that affects both male and female maize flowering time (Romero Navarro et al. 2017). Notably, BX10 and BX11, which encode the same enzymatic activity as BX12, and are located in the same region of maize chromosome 1, had no detectable effect on flowering time in this analysis. As benzoxazinoids have not been identified as direct regulators of plant growth, the role of BX12 in regulating maize flowering time is likely to be indirect. For instance, the relative abundance of DIMBOA-Glc and HDMBOA-Glc may have differing effects on plant growth rate. Based on current evidence, it is not possible to determine whether the selection for BX12 inactivation during the northward movement of maize was based primarily on herbivore resistance or flowering time. Involvement of defensive metabolites such as benzoxazinoids in regulating flowering time would not be unprecedented, as natural genetic variation in Arabidopsis glucosinolate biosynthesis has been shown to affect flowering time in this species (Jensen et al. 2015).

Examination of the DIMBOA-Glc and HDMBOA-Glc data in Fig. 2 shows that the selection for benzoxazinoid profiles was not absolute. There are certainly tropical maize varieties with relatively high DIMBOA-Glc content and temperate maize varieties with relatively high HDMBOA-Glc content. Given that at least three other methyltransferases (BX10, BX11 and BX14; Fig. 1) catalyze the methylation of DIMBOA-Glc to HDMBOA-Glc, it would be surprising if natural variation in the constitutive abundance of these two benzoxazinoids were a simple bi-allelic switch based on the presence or the absence of a transposon insertion in BX12.
A Role for Benzoxazinoids in Auxin Signaling

One way in which benzoxazinoids could affect maize growth rate is through their effect on plant developmental hormones. Similar to indole glucosinolate breakdown products in Arabidopsis (Katzy et al. 2015), benzoxazinoids and their breakdown products have been associated with regulation of auxin signaling. DMBOA (6,7-dimethoxy-2-benzoxazolinone) and MBOA were identified from maize leaf extracts as inhibitors of auxin binding to both membrane fractions and isolated membrane proteins (Venis and Watson 1978). In this binding competition assay, DMBOA was approximately 50-fold more effective than MBOA. The non-methoxylated BOA breakdown product was only weakly active. Subsequent experiments with oat, timothy grass, amaranth and pea showed that exogenous addition of DMBOA and/or MBOA affects auxin-induced growth (Hasegawa et al. 1992, Gonzalez and Rojas 1999, Nakajima et al. 2001).

Light-induced curvature, despite an even auxin distribution, has been observed in several plant species, leading to the hypothesis that this effect is caused by the induced production of an auxin inhibitor. Experiments with maize seedlings suggest that DMBOA and/or its breakdown products could constitute such inhibitors. In response to blue light illumination from one side of the plant, β-glucosidase activity and the production of DMBOA and MBOA was increased on that side (Hasegawa et al. 2004). Chemical inhibition of the β-glucosidase activity in maize plants decreased: (i) blue light-induced DMBOA and MBOA accumulation; (ii) the phototropic curvature; and (iii) the growth inhibition that was induced by the blue light treatment (Jabeen et al. 2006). Although these results suggest that benzoxazinoids are the sought-after auxin inhibitors that regulate maize phototropic growth, there are likely to be many glucosidases inhibited in these assays and the effect on coleoptile growth could be incidental. Additionally, it has been proposed that H₂O₂, which is released by DMBOA production on the illuminated side of the maize coleoptile, could contribute to the reduced growth and curvature (Jabeen et al. 2007). Further research, in particular genetic mapping of natural variation in maize auxin responses and experiments with maize mutants that have altered benzoxazinoid metabolism, will be required to confirm the hypothesis that benzoxazinoids and/or their breakdown products have direct and/or indirect effects on plant hormone signaling.

Iron and Aluminum Binding by Benzoxazinoids

The metal-chelating properties of benzoxazinoids may provide additional functions that are not directly related to plant defense. Many grasses secrete phytosiderophores into the rhizosphere to produce phytosiderophore–Fe(III) complexes that can be taken up by the plants via active transport systems (Tsednee et al. 2012). It has been proposed that benzoxazinoids, which are secreted into the rhizosphere and strongly bind iron (Tipton and Buell 1970), can also function as phytosiderophores. Not only aglucones but also the glucosides DIMBOA-Glc and DiBOA-Glc form complexes with Fe(III) (Tipton and Buell 1970, Farkas et al. 1998). Maize roots contain more DIMBOA-Glc and secrete more DIMBOA with increasing iron concentration in the medium (Petho 1992, Petho 2002), and iron complexes are taken up by maize roots in proportion to the iron needs of the plant and are transported to the shoot (Petho 1993, Petho 2002). Interestingly, not only maize roots, but also roots of oat and rice, which do not produce benzoxazinoids, take up these chelated complexes. Available bx1 and bx2 mutations in the W22 maize genetic background, which produce little or no benzoxazinoids (Betsiashvili et al. 2015, Tzin et al. 2017), could be used to investigate the in vivo role of these metabolites in iron uptake by maize.

It has been proposed that chelation of Al(III) by benzoxazinoids can lead to enhanced aluminum tolerance (Poschenrieder et al. 2005). Addition of 0.5 mM DIMBOA to a 50 μM AlCl₃ solution provided some protection against root toxicity, suggesting that naturally secreted DIMBOA might also help to protect maize root in aluminum-containing soils. However, in our hands, bx1 and bx2 transposon insertion mutations in maize inbred line W22 did not affect aluminum tolerance in an established toxicity assay (Famoso et al. 2010). At a concentration of 27 μM AlCl₃, neither bx1 nor bx2 root growth was further inhibited compared with the wild-type W22. Despite variation in the root benzoxazinoid content of different maize inbred lines, QTL mapping of aluminum tolerance has not identified known benzoxazinoid-related loci (Clark et al. 2013, Guimarães et al. 2014). However, it is not possible to rule out that as yet unknown regulatory loci or transporters determine the level of chelation of aluminum by benzoxazinoids released into the rhizosphere. Thus, the question of whether benzoxazinoids contribute to aluminum tolerance in maize remains unanswered.

Future Prospects

The newly accessible genetic analysis of maize natural variation will undoubtedly have the biggest impact on benzoxazinoid research in the next few years. With the sequencing of several maize genomes and high-density genotyping of hundreds of inbred lines, research on maize benzoxazinoid metabolism is entering a new era. It is now possible to map the underlying genetic variation in biosynthetic pathways with unprecedented detail. For instance, a GWAS using the data from Supplementary Table S1, in combination with publicly available molecular marker data (www.panzea.org), genetically maps BX12 to the single-gene level as a causative locus for the relative abundance of DIMBOA-Glc and HDMBOA-Glc in maize seedling leaves. It is likely that other missing biosynthetic genes of the benzoxazinoid pathway will be found in short order by association mapping. Analysis of gene expression QTLs will undoubtedly uncover previously unknown cis- and trans-regulatory loci that influence the developmental and defense-induced accumulation of benzoxazinoids. Additionally, it is likely that intra- and intercellular transporters of benzoxazinoids will be discovered through natural variation in their function. Beyond the discovery of the genetic basis of natural variation in maize
benzoxazinoid biosynthesis, further research will lead to the discovery of specific functions of this natural variation, not only in defense against diverse pests and pathogen, but also in other areas of maize metabolism.

**Supplementary Data**

Supplementary data are available at PCP online.

**Funding**

This work was supported by the US National Science Foundation [award IOS-1339267 to G.J.], the Binational Agricultural Research & Development Fund (award US-4846-15C to G.J.), and the Deutsche Forschungsgemeinschaft [a fellowship to A.R.].

**Acknowledgements**

We thank Edward Buckler for providing maize seeds, and Kevin Ahern for growing the plants used for benzoxazinoid assays.

**Disclosures**

The authors have no conflicts of interest to declare.

**References**


**Indole-3-Glycerolphosphate, a Branchpoint for Synthesis of Defense Compounds from the Tryptophan Biosynthetic Pathway in Maize**

A. Richter, M. Mirzaei, Adrian Powell, Navid Movahed, Georg Jander

**Introduction**

Grasses, including maize, produce a wide variety of constitutive and inducible defense compounds to protect themselves against herbivores and pathogens. Benzoxazinoids, a class of indole-derived antifeedant and insecticidal compounds in maize, provide resistance to beet armyworm (*Spodoptera exigua*), European corn borer (*Ostrinia nubilalis*), corn leaf aphid (*Rhopalosiphum maydis*), two-spotted spider mite (*Tetranychus urticae*), and other arthropod pests (McMullen et al., 2009; Meihls et al., 2012; Meihls et al., 2013; Tzin et al., 2017; Bui et al., 2018). The core benzoxazinoid biosynthesis pathway is well-studied and involves seven BX enzymes (BX1-BX5, BX8, and BX9) that catalyze the formation of 2,4-dihydroxy-1,4-benzoxazin-3-one glucoside (DIBOA-Glc) from indole-3-glycerolphosphate (IGP) (Figure 1; Frey et al., 1997; Niemeyer, 2009). The first reaction in the pathway is catalyzed by BX1, an indole-producing indole-3-glycerolphosphate lyase. Based on global coexpression network analysis, Wisecaver et al. (2017) proposed that the committed step in the benzoxazinoid biosynthesis pathway may be upstream of BX1. *IGPS1* (GRMZM2G106950), a maize gene that is co-regulated with BX genes, was predicted to encode indole-3-glycerolphosphate synthase (IGPS), which catalyzes the ring closure of 1-(2-carboxyphenylamino)-l-deoxyribulose 5’-phosphate (CdRP) into IGP (Figure 1). Maize contains two additional predicted IGPS genes, *IGPS2* (GRMZM2G169516) and *IGPS3* (GRMZM2G145870), but neither of these were co-expressed with benzoxazinoid pathway genes in maize inbred line B73 (Wisecaver et al., 2017).

Two IGPS genes, AT2G04400 and AT5G48220, have been identified in *Arabidopsis thaliana* (Arabidopsis; Li et al., 1995a). AT2G04400 enzymatic activity was verified by complementing an *Escherichia coli* mutant strain containing a missense mutation in the *trpC* gene, which encodes a bifunctional IGPS - *N*-5′-phosphoribosyl)-anthranilate isomerase (PRAI) enzyme (Yanofsky et al., 1971; Li et al., 1995a). Based on the complementation assays using *E. coli* mutants deficient in either the IGPS or PRAI activity of TrpC, it appears that plants only have monofunctional IGPS enzymes (Li et al., 1995a; Li et al., 1995b). In turn, IGP is a precursor for indole and numerous other indole-derived plant metabolites, including tryptophan,
indole alkaloids, glucosinolates, benzoxazinoids, and auxin (Rose and Last, 1994; Frey et al., 2009). Arabidopsis has two confirmed enzymes, TSA1 (tryptophan synthase alpha subunit; AT3G54640) and INS (indole synthase; AT4G02610), that convert IGP into indole. Whereas TSA is involved in the tryptophan biosynthesis, as part of a complex with TSB (tryptophan synthase beta subunit AT5G54810), INS catalyzes indole synthesis independent of TSB (Zhang et al., 2008). Unlike TSA, Arabidopsis INS does not have a chloroplast targeting sequence. Thus, the pathways for the biosynthesis tryptophan and tryptophan-independent indole-containing metabolites may be separated by metabolic channeling in Arabidopsis.

Three maize genes, *BX1* (benzoxazineless 1), *TSA* (tryptophan synthase a subunit) and *IGL* (indole glycerolphosphate lyase), encode enzymes that convert IGP to indole (Figure 1). Those genes can be differentiated by both their expression levels and their relative contributions to the biosynthesis of tryptophan, benzoxazinoids, or free indole. Whereas *BX1* is constitutively expressed in young maize seedlings, *IGL* expression is upregulated in response to insect feeding, consistent with the observed herbivory-induced increase in maize indole emission (Frey et al., 2000). Mutations in *IGL* but not *BX1* lead to an indole-deficient phenotype (Ahmad et al., 2011). Conversely, only very low amounts of benzoxazinoids can be detected in *bx1* mutant lines (Hamilton, 1964; Melanson et al., 1997; Betsiashvili et al., 2015), indicating that the indole produced by *IGL* and/or *TSA* is not easily incorporated into benzoxazinoids. Although the benzoxazineless phenotype is not rescued by *IGL* expression, application of indole to *bx1* mutant maize plants leads to a full complementation (Ahmad et al., 2011). To date, no maize *TSA* mutation has been reported, but double mutations in the tryptophan synthase ß-subunit genes, *TSB1* and *TSB2*, lead to a seedling lethal phenotype, with high indole accumulation (Wright et al., 1992). Consistent with co-expression studies (Wisecaver et al., 2017), neither *BX1* nor *IGL* mutations significantly affect tryptophan biosynthesis, suggesting that there are three mostly independent pools of indole synthesized by three different enzymes in maize (Figure 1).

Similar to IGP catabolism by the *BX1*, *TSA*, and *IGL* protein products, IGP synthesis was predicted to be encoded by three maize genes, *IGPS1*, *IGPS2* and *IGPS3*. Here we report experiments to confirm the IGPS enzymatic function and determine whether they represent an earlier branch point in the indole pathway. Bimolecular fluorescence complementation (BIFC) assays show differences in the interaction profiles. For instance, whereas IGPS1 strongly interacts with *BX1* and IGL, but not with TSA, IGPS2 has a stronger interaction with TSA.
However, analysis of IGPS mutant plants suggests that the metabolic channeling of IGP is not as strong as that of indole produced by BX1, TSA, and IGL.

**Results**

**Maize has three predicted indole-3-glycerol phosphate synthase genes**

Amino acid sequence comparisons showed greater identity between maize IGPS1 and IGPS3 (64%) than when comparing IGPS1 vs IGPS2 (52%) and IGPS3 vs IGPS2 (51%). Relative to the other two maize enzymes, IGPS2 is more similar to Arabidopsis indole-3-glycerolphosphate synthases (AT2G04400 and AT5G48220; Figure 2A), which contribute to the tryptophan biosynthesis. Together, IGPS enzymes from maize, Arabidopsis and *E.coli* form a monophyletic clade. Similarly, the indole-3-glycerol-phosphate lyases maize (TSA, BX1 and IGL), Arabidopsis (AtTSA; AT3G54640), and *E.coli* (TrpA) cluster together (Figure 2A). A third clade in the dendrogram, includes enzymes responsible for the β-reaction of tryptophan biosynthesis, the condensation of indole and serine to form tryptophan, from maize (TSB1 and TSB2), *E.coli* (TrpB), and Arabidopsis (AtTSB1; AT5G54810).

The IGPS enzyme family is known for catalyzing a carbon-carbon ring forming reaction in which the C atoms become covalently bonded (Wierenga, 2001). The three predicted maize IGPS proteins have the characteristic protein domains of this enzyme class, including a DRE-TIM (D-R-E, active site motif; TIM = triosephosphate isomerase) metallolysase domain, which includes a catalytic site with three invariant aspartate, arginine, and glutamate residues (Casey et al., 2014). It is hypothesized that the conserved arginine in the D-R-E domain stabilizes and enolates intermediates in their respective reactions. The DRE-TIM metallolysase domain includes a TIM phosphate binding domain, which is a conserved protein fold, consisting of eight α-helices and eight parallel β-strands that alternate along the peptide backbone (Farber and Petsko, 1990; Brändén, 1991; Wilmanns et al., 1991). Additionally, the IGPS enzymes contain binding sites for indole, phosphate, ribulose/trilose, and 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose-5-phosphate.

**IGPS1, IGPS2 and IGPS3 encode active indole-3-glycerol phosphate synthases**

The coding sequences of maize *IGPS1, IGPS2,* and *IGPS3* were cloned into *E. coli* plasmid vectors for expression in the *E. coli* trpC9800’ strain (Yanofsky et al., 1971), which is deficient
in the IGPS activity of TrpC, to determine whether they complement the tryptophan auxotrophic phenotype. Whereas the empty vector control grew only when tryptophan was supplemented to the minimal medium, all the three maize IGPS genes were able to complement the tryptophan-auxotrophic phenotype (Figure 2B). Bacteria transformed with IGPS1, IGPS2 and IGPS3 grew equally well with no significant differences in the growth curves. After a 1.5 h the lag phase, cultures entered into the log phase and reached the stationary phase after ~ 7 h of culture shaking.

**IGPS1 and IGPS2 are differentially expressed after herbivory in B73 and W22**

Previous experiments showed that 24 h of *S. exigua* caterpillar feeding on maize inbred line B73 resulted in a 100-fold induction of IGPS1 expression (p < 0.001) and a 5-fold response of IGPS3 (p < 0.001), but no significant changes in IGPS2 (Tzin et al., 2017; Wisecaver et al., 2017). To determine whether the response is similar in the W22 inbred line, we infested maize seedlings with *S. exigua* for 24 h and measured the transcript levels by quantitative real time PCR (qRT-PCR). Similar to the results with inbred line B73, this showed significant induction of IGPS1 (p < 0.001) and IGPS3 (p < 0.001) transcripts (Fig. 3A) and almost no transcriptional changes for IGPS2. Methyl jasmonate treatment, which regulates many insect defense responses in plants, also increased IGPS1 and IGPS3 (Figure 3B). Since Benzoxazinoids also contribute to aphid resistance (Ahmad et al., 2011; Meihls et al., 2013), we measure IGPS expression in a time course with 0 h, 2 h, 8 h and 24 h of aphid treatment and used it for 3’RNAseq analysis. Interestingly, IGPS1 showed a significant response only at 8 h in W22 (Fig. 3C), which is comparable to the results presented in Wisecaver et al. (2017), showing a significant induction 4 h after *R. maidis* infestation of inbred line B73. Although fold change of IGPS3 expression relative to controls in response to *R. maidis* feeding is small (< 2; p < 0.05), the Illumina sequencing read counts of IGPS3 (in average: 188) in W22 plants were higher than those of IGPS1 (average: 2.5) and IGPS2 (average: 20.8), even in the absence of insect feeding. This is different from inbred line B73, which showed an average of 30 read counts for IGPS1; 17.6 read counts for IGPS3 and 12.1 read counts for IGPS2 (Tzin et al., 2015). Given the fact that benzoxazinoids are both constitutively synthesized and inducible by wounding, both IGPS1 and IGPS3 could play an important role in in W22 benzoxazinoid biosynthesis. In contrast to IGPS1
and IGP3, IGP2 expression is unchanged or even slightly inhibited by insect feeding on both W22 and B73 (Figure 3; Tzin et al., 2015; Tzin et al., 2017).

Transposon insertions in IGP1 and IGP3 reduce benzoxazinoid levels

A search of publicly available transposon insertions identified mu-02540, a likely IGP1 knockout line with a Mu insertion in the second exon (Chr.7: 82883253 bp to 82883261 bp; maize W22 version 2). We used quantitative reverse transcriptase-PCR (QRT-PCR) to compare homozygous, heterozygous, and wildtype plants after methyl jasmonate induction (Fig. 4A). Heterozygous plants showed lower expression levels than wildtype and no IGP1 transcripts could be amplified from the homozygous plants. Expression of IGP3 but not IGP2 was significantly increased in the igp1 mutant line (Fig. 4B). On average, the igp1/igp1 mutant plants had significant lower levels of DIMBOA, HMBOA, DIMBOA-Glc and HM2BOA-Glc than wildtype plants. However, no significant changes could be observed between heterozygous plants and wildtype (Fig. 4C). By contrast, there were no significant changes in the tryptophan content in igp1 mutant plants relative to wildtype (Fig 4D).

A Mu-Illumina knockout mutation for IGP3 was used to investigate the involvement of IGP3 in the benzoxazinoid biosynthesis pathway. Plants were induced with methyl jasmonate to activate the IGP3 transcription, and qRT-PCR was used to verify homozygous mutants and wildtype plants. IGP3 expression was significantly reduced (p< 0.001) in the mutant line, and the other two IGP genes showed no significant change in expression levels (Figure 5A). Tryptophan abundance was unaffected by the igp3 mutations (Figure 5B). Abundance of all measured benzoxazinoids was decreased (Figure 5C), though less so than in the igp1 mutant line.

The performance of S. exigua is effected by IGP1

Spodoptera exigua, a generalist lepidopteran herbivore, grows better on bx1::Ds and bx2::Ds mutant plants, which are almost completely devoid of benzoxazinoids, than on wild-type W22 maize (Tzin et al., 2017). We conducted bioassays to determine whether the benzoxazinoid decrease in the igp1 mutant (Figure 4C) also affects S. exigua caterpillar growth. Consistent with the defensive function of benzoxazinoids, there was a significant increase in caterpillar
body mass, with an average weight of 20 mg on the mutants compared to 13.6 mg average weight on wildtype plants after nine days of feeding (p < 0.05, Fig. 4E).

**IGPSs enzymes are localized in maize chloroplasts**

In addition to gene expression, another factor that can influence the function of IGPS enzymes is their sub-cellular localization. TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/), which predicts subcellular localization and cleavage sites for the signal peptides, indicated a chloroplast location for all three maize IGPS proteins. To confirm this predicted enzyme localization, we cloned the *IGPS* open reading frames (B73-alleles) into the vector pENSg, containing an N-terminal yellow fluorescent protein (YFP) tag, as well as into pEXSG with a C-terminal YFP tag. These constructs were transformed into maize protoplasts and analyzed by confocal microscopy. Only clones containing the C-terminal tag showed a positive YFP signal after excitation with a 514 nm laser. Protoplast transformation of *pEXSG::YFP-IGPS1*, *pEXSG::YFP-IGPS2* and *pEXSG::YFP-IGPS3*, revealed similar YFP signals in maize chloroplasts for all the constructs (Fig 6A). The YFP signals overlap with the red chlorophyll autofluorescence of the plastids. In addition to the IGPS proteins, we determined the localization of BX1 and BX2, the first two enzymes of the benzoxazinoid biosynthesis pathway (Figure 1). Similar to the IGPS proteins, the intracellular localization of *pENSG::YFP-BX1* is in the plastids of transformed maize protoplasts (Fig. 6A). By contrast, BX2 carrying an N-terminal mcherry-tag was co-localized with an *ER:CFP*-marker (Fig. 6B).

**IGPS1 enzyme strongly interacts with BX1, IGL and IGPS2**

To determine whether the three IGPS enzymes interact with BX1, TSA and IGL, we performed BIFC interaction studies for all nine possible two-protein combinations. IGPS1, IGPS2 and IGPS3 were cloned into p2YC, so that their C-termini are fused to a C-terminal YFP fragment (amino acids 158-238 aa). The open reading frames for BX1, TSA and IGL were cloned into p2YN, fusing an N-terminal YFP fragment (amino acids 1-159) at the C-termini. All constructs were transformed into *Agrobacterium tumefaciens* and infiltrated into *Nicotiana benthamiana* plants. Two days after infiltration, protoplasts from the infiltrated leaf area were isolated and
analyzed under the confocal microscope. Strong YFP signals were localized in the chloroplasts for the combinations of IGPS1-C/Bx1-N and IGPS1-C/IGL-N (Fig. 7A and B). On the other hand, IGPS1 did not interact with TSA. Opposite results were detected with IGPS2, showing a strong interaction of IGPS2 with TSA, but a weaker signals with the other two enzymes, BX1 and IGL. IGPS3 consistently interacted with all three tryptophan synthase α-subunit partners. In addition, we tested whether IGPS1 is able to interact with the other two IGPSs and therefore cloned the IGPS2 and IGPS3 cDNA sequences into p2YN. An interaction with IGPS3 could be detected, suggesting a possible complex which could bind to the BX1 and IGL proteins. However, no interaction signal was detected for the IGPS1 - IGPS2 combination.

**Discussion**

**Maize contains three active IGPS enzymes channeling IGP into tryptophan, free indole and benzoxazinoid pathways**

Both complementation an *E. coli trpC* mutant and analysis of maize transposon insertion lines indicate that *IGPS1, IGPS2, and IGPS3*. Previously, genes belonging to this plant enzyme family have been characterized only in Arabidopsis (Li et al., 1995a), though indole-3-glycerolphosphate synthase activity also was verified in a wheat tissue extract (Singh and Widholm, 1974). Phylogenetic analysis revealed the greatest similarity between IGPS1 and IGPS3. Meanwhile, IGPS2 is clustered with both Arabidopsis IGP synthases. Arabidopsis does not produce benzoxazinoids and therefore has benzoxazinoid biosynthesis genes. The clustering of IGPS2 with the two Arabidopsis IGP synthases, suggests that IGPS2 may involved in the tryptophan biosynthesis rather than the production of defensive secondary metabolites.

**Indole-3-glycerolphosphate synthases do not lead to metabolic channeling**

Our results are consistent with the observations of Wisecaver et al. (2017), who used global co-expression network analysis to propose that *IGPS1* is part of the benzoxazinoid biosynthesis pathway in maize. Both B73 and W22 inbred lines show a greater than 100-fold increase of IGPS1 transcripts in response to caterpillar feeding (Figure 3A). Since *IGPS1* expression is also induced by *S. exigua* caterpillar feeding, we measured caterpillar performance on *igps1/igps1* versus wildtype W22 plants, showing significantly higher caterpillar mass on the mutant line (Figure 4E). Phloem-feeding aphids cause minimal wounding, likely explaining the relatively
low induction of IGPS1 by aphids (6-fold, Figure 3C). Indole emission is another defense reaction that is highly inducible after elicitor induction of inbred lines W22 and CI31A (Frey, 2000; Frey 2003). This suggests that, depending on the inbred line being tested, IGPS1 may also contribute to free indole biosynthesis in maize.

In contrast to IGPS1, IGPS2 doesn’t show any significant changes and IGPS3 expression increases only ~4 times in response to caterpillar feeding, aphid feeding, and jasmonic acid treatment (Figure 3). This lower induction does not preclude the involvement of IGPS3 in plant defense, which was confirmed by measuring benzoxazinoids in the igps3 mutant line, which has significantly lower benzoxazinoid levels.

We hypothesize that IGPS2 plays a role in tryptophan biosynthesis. In our studies we weren’t able to propagate the igps2 mutant line for further crosses. Interestingly, the amount of tryptophan remained unchanged in both igps1 and igps3 mutant lines, suggesting that these genes are not involved in the tryptophan biosynthesis. Again, this question could be addressed by creating a double mutant. Whereas a single knockout of igps1 or igps2 did not lead to a total loss of BXs, a single mutation in bxl or bx2 on the other hand causes the generation of very little amounts of benzoxazinoids after herbivory (Tzin et al., 2017). Thus, although IGPS1 and IGPS3 are involved in benzoxazinoid biosynthesis, they are less pathway-specific than BX1 and BX2 and also lead to the formation of other indole metabolites.

**IGPS enzyme family interacts with downstream enzymes BX1, IGL and TSA**

The regulation of defense compounds by different expression patterns of the IGPS family is only one possible mechanism of benzoxazinoid biosynthesis regulation. Another aspect to look at is the interaction ability of the IGPS with downstream enzymes. Since all three maize IGPS proteins are co-localized in the chloroplasts with IGL, BX1 and TSA, there is a possibility of specific complex formation. Whereas IGPS3 accepts all three possible binding partners, the other two enzymes seems to be more specific. IGPS1 shows interaction specifically with IGL and BX1, but not TSA, and IGPS2 strongly interacts with TSA and weakly with BX1 and IGL. These results support the idea of IGPS1 and IGPS3 providing IGP for the defense pathways and IGPS2 for tryptophan synthesis. A channeling of IGP to specific enzymes in downstream pathways seems plausible, since similar effects have been seen with BX1 and IGL. A bxl knockout phenotype cannot be complemented by the activity of IGL or TSA, only by
supplementing free indole. The same has been reported for the igl mutant, where the phenotype cannot be rescued by the enzyme activity of BX1 or TSA. One explanation for this phenotype, is a direct channeling of indole from one enzyme to the next in each branch of the pathway (Figure 1). However, in the case of IGP, the channeling to tryptophan and specialized metabolism, respectively, may be less precise.

Materials and methods

Plant material and MeJA treatment

Maize (Zea mays L.) seeds from the inbred lines B73, W22, and transposon insertion lines mu-02540 line and mu-Illumina 228854.6 were germinated in a Petri dish covered with water for three days, and were then transferred into plastic pots filled with a soil mix (0.16 m3 Metro-Mix 360 from Scotts, Marysville, OH, USA; 0.45 kg finely ground lime; 0.45 kg Peters Unimix from Griffin Greenhouse Supplies, Auburn, NY, USA; 68 kg Turface MVP from Banfield-Baker Corp., Horseheads, NY, USA; 23 kg coarse quartz sand, and 0.018 m3 pasteurized field soil). Plants were grown for about 2 weeks in climate-controlled chambers (16 h light:8 h dark cycle, 180 mmol photons m⁻² s⁻¹ light intensity at constant 23 °C and 60% humidity). For methyl jasmonate (MeJA) induction experiments, the third leaf of each plant was cut off and incubated in 2 mL tap water with 250 µM MeJA (company) for 24 h, after which the leaves were snap-frozen in liquid nitrogen and stored at -80 °C until they were used for assays.

Aphid time course assay

The experiment was performed based on Tzin et al (2015), but harvesting samples at different time points: 0, 2, 8, and 24 h. The control plants received empty cages without aphids for 24 h and were harvested and frozen in liquid N₂ at the same time as the experimental samples.

Total RNA isolation and cDNA synthesis

Harvested leaf material was frozen in liquid nitrogen, and fine powdered for RNA isolation with on-column DNA digestion, using the SV Total RNA isolation kit (Promega, Madison, WI, USA). The total RNA concentration was measured with a NanoDrop instrument (2000c;
ThermoFisher Scientific Inc., Waltham, MA, USA), and 500 ng total RNA was incorporated for first-strand cDNA synthesis using the M-MLV reverse transcriptase kit (TaKaRa Bio USA, Mountain View, CA, USA), and the library was used as a template for qRT-PCR analysis. Thereafter, 5 µl SYBR Green Mix, 1 µl gene-specific forward primer, 1 µl gene-specific reverse primer, 0.5 µl template, and 2.5 µl PCR-grade water were mixed. The PCR reaction was performed in an Applied Biosystems instrument with an initial incubation at 95°C for 10 min. The following cycle was repeated 40 times: 95°C for 30 s, 60°C for 15 s, and 72°C for 30 s. Maize adenine phosphate transferase 1 (APT1) was used as a reference gene. For each analyzed gene, a cDNA pool from all plants was diluted from 1 times to 80 times to generate a standard curve. The ∆CT (Δ-threshold cycle) for each gene was calculated relative to the reference gene. For the CT values an average from at least three biological replicates was calculated.

**Transcriptome sequencing and RNAseq data analysis**

The transcriptome of the W22 time course experiment was sequenced using the 3´RNAseq method (Kremling et al., 2018). RNA was isolated as described above, with a total of 8 biological replicates for each treatment. The purity of all RNA samples were proofed with a NanoDrop2000 instrument (Thermo Scientific). The 3´RNA-seq libraries were prepared from 6 µg total RNA at the Cornell Genomics facility (http://www.biotech.cornell.edu/brc/genomics-facility; Kremling et al. 2018). To remove the first 12 bp from each read Trimmomatic version 0.36 was used. STARaligner v.2.4.2 was used to align the reads against the W22 genome (ZmW22-REFERENCE-NRGENE-2.0, ftp://ftp.ncbi.nlm.nih.gov/genomes/genbank/plant/Zea_mays/latest_assembly_versions/GCA_001644905.2_Zm-W22-REFERENCE-NRGENE-2.0), with a maximum of one location to map to (–outFilterMultimapNmax 1). Another step in the pre-processing was increasing the region size of each gene by 500 bp to decrease the number of genes mapped to a region with no feature from > 120000 to < 60000. For this, we used the slop command from bedtools (website) with −r 500 (−r the number of base pairs to add to the end coordinate). The program Tablet (website) was used to have a look at the genome with the anotated genes for each chromosome. HTSeq v.0.6.1 were used to obtain gene-level counts from the resulting BAM files (htseq-count -r name -s yes -a 10 -f bam).
Isolation and cloning of cDNAs

To isolate the open reading frames of igps1, igps2, igps3, tsa, igl, bx1 and bx2 from B73 cDNA, specific primers were created based on the B73_v4 maize genome. For PCR amplification the iproof high fidelity polymerase from Biorad was used. The fragments were cloned into the specific vectors mentioned for each experiment and sequenced for their accuracy.

Complementation assay in E.coli trp- mutant

The open reading frames of igps1, igps2 and igps3 were cloned into the bacterial expression vector pASKIBA37+ (IBAGmbH, Göttingen, Germany). To amplify these genes, special primers with a BsaI restriction site were created with the “Primer D´Signer” (IBA GmbH) (Supplementary Table S1). After sequencing the constructs, plasmids were used for complementation assays in the E.coli tryptophan mutant background. The plasmids were transformed into the competent mutant strain trpC9800- and plated out on Ampicillin (Amp) LB-agar plates. To test their activity in the mutant background, a starter overnight culture of 2 mL LB-Amp medium was prepared for each construct plus empty vector as a negative control. The next morning, the starter culture was used to inoculate 5 mL of tryptophan-deficient Vogel and Bronner (1956) minimal medium containing 2 % dextrose and ampicillin (100 ug/mL). For positive controls L-tryptophan (20 ug/mL) was added. All cultures started with an OD600 = 0.2 in the time-course experiment. OD600 was measured six times over 11 hours, until the bacteria reached the stationary phase.

Phylogenetic tree

The dendrogram analysis was conducted as described in Richter et al. (2016). BX3 was used as the outgroup.

Untargeted Metabolite analysis via LC-MS and data processing
Frozen, powdered leaf material was weighed (~100 mg), three volumes of 100% methanol were added to each sample, and samples were incubated for 45 min shaking at 4 °C. After a 10 min centrifugation step (15,000 x g) the samples were filtered in a 96-well filter plate. All samples were analyzed on a Sigma Supelco reverse phase C18 column on a Dionex 3000 Ultimate UPLC-diode array detector system coupled to a Thermo Q Exactive mass spectrometer. For the two mobile phase solvents, water (Solvent A) and acetonitrile (Solvent B), both acidified with 0.1% formic acid, were used.

The reverse-phase liquid chromatography was performed using a Dionex Ultimate 3000 Series LC system (HPG-3400 RS High-Pressure pump, TCC-3000RS column compartment, WPS-3000TRS autosampler) controlled by Chromeleon Software (Thermo Fisher Scientific) and coupled to OrbitrapQ-Exactive mass spectrometer controlled by the Xcalibur software (Thermo Fisher Scientific). A Titan™ C18 UHPLC Column, 1.9 μm 100 mm × 2.1 mm at 40°C and the flow rate of 0.5 mL/min of mobiles phases A (H2O:0.1% formic acid) and B (AcN:0.1% formic acid) was used for the separation of target molecular features. The gradient starting condition was 0% B at 0 min, rising to 95% B in 15 min, then was held for 0.5 min, and followed by 0.5 min re-equilibration to the starting condition. A heated electrospray ionization source (HESI-II) in negative mode was used for the ionization with parameters of spray voltage at 3.5 kV, the capillary temperature at 380°C, sheath gas and auxiliary gas flow at 60 and 20 arbitrary units respectively, S-Lens RF Level 50, and probe heater temperature 400°C. The data were acquired in the m/z range of 100–900, 140,000 FWHM resolution (at m/z 200), AGC target 3e6, maximum injection time of 200 ms, in profile mode. Metabolite quantification was estimated with ThermoScientific Xcalibur™ Version 4.1.31.9 (Quan/Processing). For data analysis the raw mass spectrometry files were converted to mzxml formats using the MSConvert tool (Chamber et al., 2012). The raw mzxml files were further processed using the XCMS (http://metlin.scripps.edu/download/, Tautenhahn et al., 2008) and the CAMERA software packages for R (http://www.bioconductor.org/packages/release/bioc/html/CAMERA.html). Data were analyzed using the XCMS-CAMERA mass scan data processing pipeline (Tautenhahn et al., 2008; Benton et al., 2010; Kuhl et al., 2012).
Protoplast isolation and transformation

Maize mesophyll protoplasts were isolated based on Richter et al., 2016 with few modifications. After washing the protoplasts, the cells were transferred into the wash buffer instead of the MMG buffer, with a final concentration of ~ 1-2 x 10⁶/mL. The protocol was proceeded as described in Richter et al. (2016).

Bimolecular Fluorescent Complementation (BiFC) Localization and Co-Localization Assays via Confocal Laser Scanning Microscopy

The open reading frames of IGPS1, IGPS2, IGPS3, BX1, BX2, TSA and IGL were amplified without stop codon and cloned into pDONR 201 by BP clonase recombination (Invitrogen), before it was transferred in a second recombination reaction (LR, Invitrogen) into the vector YFP-pENSg (Dissmeyer et al, 2017). To isolate high concentrations of each plasmid a Midi preparation (Quiagen) was performed. All the fusion proteins were transformed into maize mesophyll protoplasts as described above. After incubating maize protoplasts overnight, they were analyzed under the Leica TCS SP5 spectral imaging system (Leica Microsystems, Wetzler, Germany). The YFP fluorescence was excited with a 514 nm laser and the emission spectra were collected with a hybrid detector (HyD) at 523 nm to 591 nm. The autofluorescence of chloroplasts was excited with an argon laser and detected at 675 to 736 nm. Protoplasts from each transformation were imaged with a 20x water immersion objective. For co-localization of BX2 with the ER the open reading frame was cloned from pDONR 207 via LR gateway recombination. For this reaction the destination vector, pDONR, containing the 35S promoter and a plasmid with an mcherry tag were fused together. Finally, the construct 35S:mcherry:BX2 was transformed into A. tumefaciens (GV3101) and grown overnight in LB containing appropriate antibiotics. Likewise, an ER-CFP marker construct (plasmid #953, Nelson et al., 2007) was transformed into Agrobacterium as well. Agrobacterium cultures of both constructs were consequently co-infiltrated into N. benthamiana leaves, by mixing both cultures with the same OD600. CFP fluorophore excitation was performed by using a 458 nm argon laser and detected at 468 to 496 nm. For co-localization CFP and mcherry were scanned sequentially from each other. For interaction studies the open reading frames of IGPS1, IGPS2, IGPS3, TSA, IGL, and
BX1 were cloned without stop codons into the BIFC vectors p2YC and p2YN (Kong et al., 2014). In addition, IGPS1 was cloned into p2YN to study the interaction with IGPS2 and IGPS3. The constructs used for protein-protein interactions studies, were co-infiltrated into N. benthamiana leaves (Castañeda et al., 2005). To detect the YFP signal and chloroplast autofluorescence the same settings were used as described above.

**Caterpillar bioassay**

For the caterpillar assay, 11-day-old homozygous igps1 mutant and wildtype W22 plants were used. Spodoptera exigua eggs were ordered from Benzon Research (Carlisle, PA, USA) and incubated for 52 h in a 28°C incubator on an artificial diet (Beet Armyworm Diet, Southland Products Inc., Lake Village, AR, USA). On each plant one organza bag (https://www.amazon.com/gp/product/B073J4RS9C/ref=ppx_yo_dt_b_asin_title_o08_s00?ie=UTF8&psc=1) was placed, containing one second instar S. exigua caterpillar. The bags were gently closed with bop pins without wounding the plant. After one week of infestation, the caterpillars were weighed to assess differences in their performance. Statistical comparisons were performed with Microsoft Excel.

**Acknowledgements**

This research was funded by a Deutsche Forschungsgemeinschaft fellowship to AR and by Research Grant Award No. US-4846-15C from BARD, The United States - Israel Binational Agricultural Research and Development Fund, to GJ. We thank Monika Frey for helpful advice.

**References**


Brändén C-I (1991) The TIM barrel—the most frequently occurring folding motif in proteins. 
Curr Opin Struct Biol 1: 978-983

and specialist mite herbivores induce similar defense responses in maize and barley but 
differ in susceptibility to benzoxazinoids. Front Plant Sci 9: 1222

Casey AK, Hicks MA, Johnson JL, Babbitt PC, Frantom PA (2014) Mechanistic and 
bioinformatic investigation of a conserved active site helix in alpha-isopropylmalate 
synthase from Mycobacterium tuberculosis, a member of the DRE-TIM metalloclyase 
superfamily. Biochemistry 53: 2915-2925

Sci 15: 228-234

Bacher A, Meeley RB, Briggs SP, Simcox K, Gierl A (1997) Analysis of a chemical 

a model for evolution of secondary metabolic pathways in plants. Phytochemistry 70: 
1645-1651

elicitor activates the gene for indole emission in maize. Proc Natl Acad Sci U S A 97: 
14801-14806

Hamilton RH (1964) A corn mutant deficient in 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-
one with an altered tolerance of atrazine. Weeds 12: 27-30

Li J, Chen S, Zhu L, Last RL (1995a) Isolation of cDNAs encoding the tryptophan pathway 
enzyme indole-3-glycerol phosphate synthase from Arabidopsis thaliana. Plant Physiol 
108: 877-878

Li J, Zhao J, Rose AB, Schmidt R, Last RL (1995b) Arabidopsis phosphoribosylanthranilate 
isomerase: molecular genetic analysis of triplicate tryptophan pathway genes. Plant Cell 
7: 447-461

McMullen M, Frey M, Degenhardt J (2009) Genetics and biochemistry of insect resistance in 
York, p 587


The maize W22 genome provides a foundation for functional genomics and transposon biology


The maize W22 inbred has served as a platform for maize genomics since its early adoption by Brink for studies of paramutation1. The inbred, developed at the Wisconsin Agricultural Experiment Station, lacked the ubiquitous suppressors of anthocyanin pigmentation present in most standard corn belt inbreds but did carry mutations in two regulatory genes controlling anthocyanin production, R1 and C1. Thus, Brink introgressed functional alleles of R1 and C1 to produce a color-converted W22 inbred (PI accession 674445). This inbred was soon adopted by Nelson, Kermicle and Dooner and their students, who led foundational studies to understand the genetics of high-protein maize (O2)’, provide the first example of imprinting1, conduct the first transposon tagging experiments in plants1 and understand the mechanisms of Activator and Dissociation10 transposition. Today, the W22 genome is the host to thousands of mobilized, indexed Mutator (Mu) and Dissociation (Ds) transposable element insertions for reverse and forward genetics studies. Annotation of the genome has been achieved using RNA-seq analysis, differential nuclease sensitivity profiling and bisulfite sequencing to map open reading frames, open chromatin sites and DNA methylation profiles, respectively. Collectively, the resources developed here integrate W22 as a community reference genome for functional genomics and provide a foundation for the maize pan-genome.

The maize W22 genome has served as a foundation for maize genomics since its early adoption by Brink for studies of paramutation1. The inbred, developed at the Wisconsin Agricultural Experiment Station, lacked the ubiquitous suppressors of anthocyanin pigmentation present in most standard corn belt inbreds but did carry mutations in two regulatory genes controlling anthocyanin production, R1 and C1. Thus, Brink introgressed functional alleles of R1 and C1 to produce a color-converted W22 inbred (PI accession 674445). This inbred was soon adopted by Nelson, Kermicle and Dooner and their students, who led foundational studies to understand the genetics of high-protein maize (O2)’, provide the first example of imprinting1, conduct the first transposon tagging experiments in plants1 and understand the mechanisms of Activator and Dissociation10 transposition. Today, the W22 genome is the host to thousands of mobilized, indexed Mutator (Mu) and Dissociation (Ds) insertions that are used extensively for functional genomics studies. Maize exhibits exceptional levels of structural variation that influences gene content and gene regulation among

1Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, MN, USA. 2USDA-ARS, Corn Insects and Crop Genetics Research Unit and Iowa State University, Department of Computer Science, Iowa State University, Ames, IA, USA. 3Boyce Thompson Institute, Ithaca, NY, USA. 4Horticultural Sciences Department, University of Florida, Gainesville, FL, USA. 5NRGene Ltd, Ness Ziona, Israel. 6Department of Biology and the UF Genetics Institute, University of Florida, Cancer & Genetics Research Center, Gainesville, FL, USA. 7Department of Biological Science, The Florida State University, Tallahassee, FL, USA. 8USDA-ARS, Holley Center for Agriculture and Health, Ithaca, NY, USA. 9Institute for Genomic Diversity, Biotechnology Building, Cornell University, Ithaca, NY, USA. 10Cold Spring Harbor Laboratory, Cold Springs Harbor, NY, USA. 11Department of Plant Biology, University of Georgia, Athens, GA, USA. 12Department of Plant Biology, Rutgers University, New Brunswick, NJ, USA. 13Waksman Institute, Rutgers University, Piscataway, NJ, USA. 14Department of Biological Science, Montclair State University, Montclair, NJ, USA. 15USDA-ARS Chemistry Research Unit, Gainesville, FL, USA. 16Department of Biochemistry, Max Planck Institute for Chemical Ecology, Jena, Germany. 17Metabologonomics, Inc., Tsuruoka, Yamagata, Japan. 18CAS-JIC Centre of Excellence for Plant and Microbial Science (CEPAMS), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. 19The State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. 20USDA-ARS, NEA Robert W. Holley Center for Agriculture and Health, Cornell University, Ithaca, NY, USA. 21Present address: College of Agronomic Sciences State Key Laboratory of Crop Biology, Shandong Agricultural University, Shandong, China. *e-mail: brutnell@gmail.com
different inbred lines\textsuperscript{14–17}. This limits the utility of the B73 reference genome\textsuperscript{18} or the PH207 genome\textsuperscript{19} for interpreting data derived for mutant alleles in a W22 genetic background. Detailed knowledge of the W22 genome structure and content is critical to fully exploit reverse genetics resources and biological processes in W22.

**W22 genome de novo assembly and validation**

The W22 genome was sequenced and assembled using deep sequencing (> 180x) of Illumina short-read sequences from libraries with a variety of insert sizes using the assembly algorithms developed by NRGene\textsuperscript{20} with the identifier Zm00004b (see Methods and Reporting Summary for details). Sequence contigs were developed based on paired-end (PE) short reads and then were put into scaffolds using PE or mate-pair (MP) links. Further ordering and orientation of scaffolds was achieved through alignment to the B73 genome assembly. The initial assembly was improved through utilization of the 10x genomics library platform\textsuperscript{21} and sequencing using the HiSeq X10 to bridge contigs and correct initial misassemblies (see Methods). This revealed 34 potential chimeric scaffolds that were broken and reassembled to generate a W22v2 genome (Table 1; see Methods). The W22 whole genome is available at NCBI under the accession GCA_001644905.2. The genome and accompanying resources are available at MaizeGDB under the name Zm-W22-REFERENCE-NRGENE-2.0 with the identifier Zm00004b.

Several approaches were used to assess the quality and completeness of the W22 genome sequence at macro and micro scales. To confirm the overall quality of the sequence assembly, we created an independent optical map based on the restriction site BspQI\textsuperscript{21} and aligned the optical map with an in silico digested version of the genome assembly. The two assemblies strongly supported each other, showing exceptional (>98%) alignment and linearity. Further confirmation of the assembly was performed using a suite of 4.4 million SNP markers validated against the maize nested association mapping panel\textsuperscript{22}. A genetic assessment of contig and scaffold ordering was conducted by aligning 4.4 million SNP markers to an initial W22 assembly scaffold revealing ~30 inconsistencies (see Methods). The optical map was used to correct the initial assembly to produce a W22v2 assembly that integrates the genetic and physical maps and provided evidence for robust assembly. The W22v2 genome has a larger number of gaps (as defined by >10Ns) than the B73v4 genome but substantially fewer gaps than the PH207 genome (Supplementary Table 1). A more granular assessment of the assembly quality and accuracy was made by comparing the W22v2 assembly to a 238-kilobase (kb) interval encompassing the W22-bz1 locus on chromosome 9 that had previously been sequenced using Sanger sequencing technology\textsuperscript{23}. The vast majority (~113 Mb) of this region aligned with >99% identity in a collinear fashion (Fig. 1). The primary difference between the W22v2 genome assembly and the earlier Sanger BAC assembly is an unfilled gap present in the W22v2 assembly that corresponds to 7,137 base pairs (bp) of a Misfit retrotransposon that is located in the midst of an 85-kb retrotransposon cluster (Fig. 1). There are three other small gaps in this region that occur within repetitive elements but the size of these gaps is similar to the BAC sequence and the order and orientation is the same as the BAC. This analysis provides further evidence for the quality of the W22v2 assembly and suggests that remaining gaps likely occur within repetitive regions.

**Annotation of genes and transposons in W22**

Genes of the W22v2 genome were annotated using a modified Maker P approach that was developed for B73v4\textsuperscript{27} together with RNA-seq from W22 tissues and full-length cDNA data (see Methods for details). A ‘working gene set’ of 57,181 genes was identified. Removal of putative transposons, low-confidence single-exon genes and potential pseudogenes resulted in a ‘filtered gene set’ of 40,789 genes. The exons of the filtered gene set account for 2.99% and the introns account for another 5.15% of the 2.2 Gb W22v2 assembly. The number of genes and distribution along the chromosomes (Fig. 2) are very similar to previous maize genome assemblies for B73\textsuperscript{4} and PH207\textsuperscript{4}. The quality and completeness of the W22 genome was assessed through a BUSCO analysis\textsuperscript{4}. Over 90% of BUSCO genes are represented by a complete single copy in the W22 genome and only 3% of BUSCO genes are missing in the W22v2 assembly (Supplementary Fig. 1). These rates are equivalent to the B73v4 genome (Supplementary Fig. 1), suggesting similar qualities of annotation completeness and quality in these genomes.

The availability of a de novo assembly for the W22 genome allows for comparisons of gene copy number, structure and splicing between W22 and B73. Previous work has found evidence for frequent copy number variation and presence/absence variation among maize inbreds\textsuperscript{13,19,25}. The B73 and W22 genomes were compared to each other and to the Sorghum BTx623 genome\textsuperscript{26} to document the presence or absence of syntenic orthologs among these three genomes (Supplementary Table 2). Over 60% of the genes are present at syntenic locations in both maize lines and Sorghum. Another 15–20% of genes are found in both B73 and W22 but are absent from the BTx623 assembly (Supplementary Table 2). The remaining genes are present in one maize line and absent in the other. The majority of these genes are present only in B73 (6,440) or W22 (8,372) and not found in the other maize inbred or in Sorghum. There are relatively fewer examples of genes that are present in Sorghum and only one maize inbred (B73, 881; W22, 1,412). The presence of locally duplicated genes was evaluated in both the W22 and B73 genome, for a range of intervening genes from zero (tandem duplicates) to twenty (Supplementary Tables 3 and 4). A subset (678) of the examples of tandem duplications are observed in both B73 and W22. However, there are many other cases of tandem duplication that are observed only in B73 (1,394) or in W22 (1,261). The tandem duplicates were classified according to their relative arrangement (Supplementary Table 5). Tandem duplicates in the same orientation (head-to-tail) are more common than genes in a convergent (tail-to-tail) or divergent (head-to-head) arrangement (Supplementary Table 5).

The presence of a tandem duplication in one genome and not in the other could reflect mechanisms to rapidly expand or contract gene copy variants and may serve as a foundation for functional diversification; for example, the terpene synthase family has variable tandem duplications that may affect function (see Supplementary Note for details).

RNA-seq data generated from several W22 tissues were used to document examples of alternative splicing in W22 and compare these alternative splicing events to those observed in B73\textsuperscript{4}. We identified 222 genes with alternative splicing, representing 49.4% of expressed multi-exon genes. The types, and frequencies, of alternative splicing events noted in W22 (Supplementary Table 6) are similar to those noted for B73\textsuperscript{4}, albeit at lower numbers, which
Articles

ARTICLES

NATURE GENETICS

Fig. 1 | Benchmarking the W22v2 assembly quality and completeness. A comparison of the W22v2 assembly with a Sanger-based assembly of a 238-kb maize BAC (Genbank: EU338354). This region contains five genes (yellow/tan symbols along line) as well as a number of transposons (colored triangles). The W22v2 genome matches the majority of the sequence at quite high identity (>99%) and the locations of gaps in the W22v2 genome (indicated by red dotted lines).

Fig. 2 | Distribution of features within the W22 genome. The relative density for a variety of features in 1 Mb bins across the maize genome is shown. The purple tracks show the density of genes, LTR transposons and TIR transposons. The relative chromatin accessibility as assessed by MNase (green) and the levels of CG, CHG and CHH methylation (red) are shown. The frequencies of Ds or Mu transposon insertions are also shown (gray).

may be due to reduced levels of RNA-seq data for W22 relative to B73. There are numerous locus-specific differences in splicing observed between B73 and W22 and these are found distributed throughout the genome (Supplementary Fig. 2). One major advantage of sequencing and annotating additional genotypes is the improved ability to analyze high-throughput sequencing data sets. To test the benefit of using a genotype-matched reference, W22 leaf RNA sequencing data (SRA: SRR1986376) was analyzed in parallel using both the W22 and the B73v4 genome references. Mapping RNA-seq reads from W22 seedling leaf tissue to the W22 genome resulted in a mapping rate of 95.7% while only 91.1% of these reads could be aligned to the B73v4 genome (Supplementary Fig. 3a). A comparison of the expression level for orthologous genes in B73 and W22 reveals that the expression level estimates vary for a number of genes (Supplementary Fig. 3b). In many cases, these changes in expression were due to altered annotation of the transscripts in W22 relative to B73 (Supplementary Fig. 3c).

The maize genome is largely composed of transposable elements (TEs)25. A structural annotation of transposons in the W22v2 genome identified over 177,000 transposons that were classified into 26,833 families (see Methods). Long terminal repeat (LTR) retrotransposons are the most prevalent type of TEs in the W22 genome with 23,144 families accounting for 64% of the genome. The 10 most abundant LTR families have 1,400–16,395 members and account for 36% of the genome. Terminal inverted repeat (TIR) DNA transposons were classified into 5 major types and account for 0.46% of the genome. There are also 23,895 helitron elements that account for 4.6% of the genome. While many of the TEs are present in intergenic space, there are numerous examples of transposons being annotated within the introns of genes. Nearly 9% of the FGS genes contain an annotated transposon, including 1,626 genes containing a TIR element and 1,864 genes containing an LTR element. A comparison of the TEs present in the B73 and W22 genome revealed high levels of variation in TE presence and copy number for both TIR and LTR families. There are 937 families of TIR elements that are found in both B73 and W22. Many other TIR families are found only in B73 (107 families) or W22 (62 families) after excluding B73 families that are present only as MITEs. These B73-specific and W22-specific TIR elements are enriched for CACTA (DTc) families (Supplementary Fig. 4a). There are 12,740 families of LTR elements annotated in both B73 and W22 and many other families specific to W22 (10,531) or B73 (11,032). The number of elements in each family is generally similar in the two genomes but there are many examples of strong copy number variation (Supplementary Fig. 4b–d). However, there are 23 LTR families and 30 TIR families with differences in copy number >100 between the two genomes including 7 with higher copy number in W22 and 26 with higher copy number in B73. The analysis of age of LTR elements (based on sequence identity of the two LTRs) in families that are inbred-specific reveals that small inbred-specific families tend to be relatively old and thus may be being purged from the genome while larger inbred specific families tend to be enriched for recent insertion events (Supplementary Fig. 4e). Even in TIR and LTR families with similar copy number we see variation for specific elements. For example, among the 100 TIR families with a single copy in each genome there are 74 examples in which the transposon is inserted into different chromosomes in the two inbred lines. A detailed analysis of native Mutator insertions in the B73 and W22 genomes (see Supplementary Note for details) revealed similar numbers in both genomes (258 in W22 and 259 in B73). However, fewer than half (125) of these insertions are shared in both lines. There are over 100 unique Mutator insertions in each line and 63% of these are located within 500 bp of an annotated gene. Among the 7,463 LTR families with a single genome.
copy in both B73 and W22 there are 863 examples in which that single element is located on different chromosomes in the two genomes. These examples of polymorphic TIR and LTR elements highlight the exceptional diversity of TE insertion events among maize genotypes.

**Mapping of transposon insertion sites using the W22 genome**

Significant efforts have been dedicated to the development of stocks with novel transposon insertions in the W22 genetic background. These include several approaches that have mobilized Ds elements and the development of a large number of Mu insertion stocks\(^7\)–\(^9\). A large number of these novel insertions have been sequence indexed to document information on the sequence flanking insertion sites. To date, the insertion sites have often been determined through mapping of flanking sequences to the B73 genome. However, a large number of insertions could not be accurately placed due to insufficient sequence identity with B73. To precisely position the 2,029 Ds and 68,866 Mu insertions that have been generated, sequences adjacent to the transposon were mapped against the W22v2 genome. This provided novel insertion sites for 12%–15% of these insertions (Fig. 3). For the 13,444 UniformMu stocks, the mapping of flanking sites to the W22 genome provided novel insertion site information for over 10,000 insertions and increased the properly placed insertions from 4.35 per line to 5.1 per line. Furthermore, the genomic distributions of Ds and Mu insertions are distinct from one another (Fig. 2), indicating complementarity of these resources for functional genomics studies. The improved mapping resolution of these insertions coupled with a better resolution of gene content and position afforded by W22v2 assembly will greatly increase the utility of the Ds and Mu reverse genetics resources. In particular, the Ds resource is designed for remobilization of elements to flanking regions of the genome, so accurate placement of insertions to gene models and local gene content is imperative for efficient mutagenesis. Similarly, better gene models will enable a more accurate prediction of phenotypic consequences of transposon insertions, particularly those that map to the 5’ and 3’ ends of a gene and thus may induce subtle mutant phenotypes.

**Ds and Mu probe different compartments of the genome**

Ds and Mu represent two different transposon superfamilies (DTA and DTM). Previous studies have suggested differences in terms of the frequency of linked transposition and insertion locations relative to genes\(^1\)–\(^3\),\(^31\). Several properties of chromatin in the W22 genome were analyzed to better understand the factors that influence the insertion sites for Ds and Mu. The context-specific DNA methylation patterns throughout the W22 genome were documented using whole-genome bisulfite sequencing (WGBS) and chromatin accessibility was assessed using MNase treatments (Figs. 2 and 4a–c).

---

**Fig. 3** Improved ability to document the location of Ds and Mu insertions using the W22 genome. The Ds (n = 2,029) and Mu (n = 68,866) insertion flanking sequences were mapped to the W22 genome. The proportion of insertion sites that mapped to the same location (black), different locations (blue) or only to W22 (red) was determined.

---

**Fig. 4** Patterns of chromatin over maize genes and transposons. a–c. The levels of CG methylation (black), CHG methylation (blue) or CHH methylation (red) over the 1kb flanking regions and maize genes (a), TIR transposons (b) or LTR transposons (c). In addition, the enrichment for open chromatin (green) based on MNase digests is shown for the same regions. d–f. Chromatin profiles for sites targeted by Ds (solid lines) or Mu (dashed lines). The relative levels of CG methylation (d) and CHG methylation (e) or open chromatin (f) are shown for the flanking regions (up to 10 kb) near Ds or Mu insertion sites.
The levels and genomic distribution of DNA methylation in W22 are similar to patterns observed in other maize inbred lines\textsuperscript{31,32}. Open chromatin and CHH methylation are substantially enriched in promoter regions and following the transcription termination sites of annotated genes (Fig. 4a). In contrast, CHG and CG methylation are relatively depleted in these regions (Fig. 4a). CG and CHG methylation are enriched within TIR and LTR transposons (Fig. 4b,c). CHH methylation is enriched at TIR elements but shows minimal enrichment at LTR elements. Chromatin accessibility does not vary substantially within or near transposons in the W22 genome (Fig. 4b,c).

The locations for novel transposon insertions can be quite sensitive to chromatin structure\textsuperscript{25}. The profiles of open chromatin and DNA methylation were assessed for the regions flanking potential Ds and Mu insertion sites (Fig. 4d–f and Supplementary Fig. 5). There are differences in the accessibility of chromatin at Mu and Ds insertion sites. The Mu insertion sites are marked by an increase in chromatin accessibility while Ds insertion sites do not vary substantially compared to flanking regions. Both Mu and Ds prefer to insert in regions with very low CG and CHG methylation. A metaprofile from bisulfite sequencing reveals that Mu insertion sites are very near regions with low methylation for flanking regions while other families are more likely to have low levels of CG and CHG methylation. The W22 genome contains a number of native TIR transposon families. The patterns of chromatin within these TIR elements likely reflects endogenous genomic information that influences chromatin state as well as the influences of the TE on neighboring chromatin. Given the differences in chromatin profiles for the insertion sites of Ds or Mu, we were interested in comparing the chromatin at these regions with the chromatin flanking pre-existing TIR insertions. We collected chromatin profiles for 2 kb of flanking regions for all TIR families with >50 elements and clustered these profiles together with the profile of chromatin flanking potential Ds and Mu insertion sites (Fig. 5). There are several quite distinct patterns of chromatin flanking TIR elements. Some families show very high methylation for flanking regions while other families are more likely to have low levels of CG and CHG methylation. The Mu and Ds potential insertion sites are outliers for the levels of CG and CHG methylation compared to the regions flanking pre-existing element insertion sites. The analysis of the methylation levels for the first 100 bp at the edge of the TE reveals that both Mu and Ds insertion sites have much lower levels of CG and CHG methylation than observed for pre-existing TIR insertion sites (Fig. 5b,c). In contrast, the levels of CHH methylation near Mu and Ds potential insertion sites are very similar to the levels of CHH methylation at pre-existing TIR insertion sites (Fig. 5d).

**Fig. 5** Chromatin profiles for regions flanking pre-existing TIR elements or novel Ds or Mu insertion sites. a. For 191 TIR families with at least 20 elements present in the W22 genome, we determined the relative level of CG and CHG methylation in 100-bp windows for the 1-kb flanking regions. These levels of methylation were used to cluster the families and a heatmap allows visualization of the patterns of DNA methylation for flanking regions. b,c. The distributions of the average level of methylation for CG (b) or CHG (c) and CHH (d) contexts in the 100 bp surrounding each TE family are shown as density plots. The levels of methylation at sites where Mu (red arrows) or Ds (black arrows) elements can insert are shown as arrows.
Since the turn of the twentieth century, maize has provided a model system for genetic research. The development of a color-converted W22 inbred enabled high-resolution mapping of numerous traits influencing anthocyanin and flavonoid accumulation and led to the adoption of W22 as the workhorse of maize genetics. In recent years, the genomics data sets generated from studies of transposon insertion alleles have been analyzed in the context of the B73 reference genome. Numerous studies have documented the widespread genetic variation among inbred genotypes of maize including SNPs, copy number variation, and indels. The availability of a high-quality assembly and annotation for W22 will increase the resolution of studies of mutant alleles in this genetic background and recent improvements in maize transformation are likely to enable higher efficiencies of W22 transformation. Thus, W22 is well positioned to serve as the maize reference inbred for functional genomics studies.

The annotation of TEs present in the W22 and B73 genome will provide new opportunities to study the sources, and effects, of TE polymorphisms among maize lines. While there are many shared transposons in B73 and W22, there are numerous examples of severe copy number and presence/absence variation of certain families. There are also many examples of polymorphic TE insertions within or near genes that could influence regulation of the gene. Locus-specific studies have provided evidence for the importance of polymorphic TE insertions that influence the regulation of nearby genes, resulting in quantitative trait loci affecting important agronomic traits. The ability to have a high structural resolution of both genomes will serve as a resource to enable detailed studies of how transposons influence gene expression among diverse maize genotypes. The knowledge of complex haplotype variation can also be used to inform pan-genome-based approaches that incorporate rearrangements and insertions/deletions from large-scale resequencing studies to augment SNP-based GWAS studies.

The availability of epigenome profiles for DNA methylation and chromatin accessibility provides an opportunity to study chromatin profiles for thousands of Mu and Ds insertion sites. The differences in these profiles suggest that Mu and Ds utilize distinct targeting mechanisms and occupy different niches of the W22 genome. Mu elements tend to insert in small windows of accessible chromatin that are near highly methylated regions. Previous work has noted the propensity for Mu elements to insert near gene promoters and these regions likely contain the chromatin profiles noted for Mu elements. In contrast, Ds frequently inserts into regions with low CG and CHG methylation that are not enriched for accessible chromatin. This is compatible with the concept of preferential insertion of Ds elements within the coding regions of maize genes. Interestingly, chromatin profiles for random sites selected within gene promoters or coding regions do not match the profiles observed for Mu or Ds sites. This suggests that Mu and Ds have specific targeting beyond simple selection for promoters or coding regions. Furthermore, the chromatin profiles at loci targeted by active Mu or Ds elements are distinct from the chromatin profiles at sites that flank inactive pre-existing insertion sites. This likely reflects the fact that many of these pre-existing element insertion sites are targeted for silencing and the chromatin properties of the elements themselves can spread into the flanking regions. It is possible that many of the pre-existing TIR families might insert into regions with chromatin profiles similar to Mu or Ds if they were active. As additional high-quality maize genome assemblies become available, comparative analysis of the maize pan-genome is likely to provide further insight into how transposons have shaped the maize genome through the course of domestication.

Received: 19 October 2017; Accepted: 17 May 2018; Published online: 30 July 2018

References


Acknowledgements

This work was supported by grants from the following: National Science Foundation NSF-MCB 0920218 to H.K.D.; NSF-MCB 1025976 to K.E.K.; NSF-IOS 1031416 to J.J.M.; NSF-IOS 1237931 to N.M.S.; NSF-IOS 1116561 to D.R.M. and K.E.K.; NSF-IOS 1314143 to T.P.B. and E.V.; NSF-IOS 1392373 and IOS-119329 to G.J.; NSF-MCB 1412218 to A.M.S. and W.B.B.; NSF IOS-1444514 to R.K.D.; NSF-IOS 1521100 to J.-C.G., K.E.K. and D.R.M.; NSF-IOS 1547787 to W.B.B.; NSF Graduate Research Fellowship and University of Florida Graduate Student Fellowship to C.G.; NSF Graduate Research Fellowship DGE-0802270 to C.G.; the United States Department of Agriculture (USDA) in kind to C.M.A., E.K.S.C., M.R.W. and J.L.P.; USDA-ARS 6036-21000-011-00D to C.T.H.; USDA-NIFA 2011-67013-24747 to N.M.S.; USDA-NIFA 2011-67003-30215 to D.R.M. and K.E.K.; USDA-NIFA 2011-67013-30082 to D.R.M. and M.S.; USDA-ARS and the NSF-IOS Plant Genome 1238014 to E.S.B.; the Enterprise Rent-A-Car Institute for Renewable Fuels at the Donald Danforth Plant Science Center to T.P.B.; and the Vasil-Monsanto Endowment to A.M.S. The use of trade names, commercial products or corporations in this publication is for the information and convenience of the reader and does not imply an official recommendation, endorsement or approval by the US Department of Agriculture or the Agricultural Research Service for any product or service to the exclusion of others that may be suitable. USDA is an equal opportunity provider and employer.

Author contributions


Competing interests

Several authors are employed by companies that provide genome sequencing and assembly support: O.B., K.B., G.B.-Z., C.G., K.E.K., M.R.W., D.S.-T. and I.S. are employed by NGene and J.S. is employed by BioNano Genomics.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0158-0.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to T.P.B.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
**Methods**

**W22 genome sequencing and assembly.** DNA sequencing. Genomic DNA was isolated from 10-day-old W22 seedling tissues and nuclear DNA was purified and fractionated from organelar DNA on a CsCl gradient (3). A genomic library of 800-bp DNA fragments was prepared using the Truseq DNA Sample Preparation Kit version 2 according to the manufacturer’s protocol (Illumina). A second shotgun library was made using the same kit from DNA template fragments size selected of ~450 bp with no PCR amplification (PCR-free). This fragment size was chosen to produce a sequencing overlap of the scaffolds to be sequenced on the HiSeq2500 v2 Rapid mode as PE sequencing 250 nucleotides per end, thus creating an opportunity to produce ‘stitched’ reads of approximately 250 bp to 480 bp in length. Multiple MP libraries were prepared with the objective to increase sequence diversity and genome coverage including three size classes (2–4 kb, 5–7 kb and 8–10 kb) using the Illumina Nextera Mate-Pair Sample Preparation Kit (Illumina). The 800-bp shotgun library and the MP libraries were sequenced on an Illumina HiSeq2500 as PE with 160 bp read lengths for each of the fragments using the SBS sequencing kit version 4. Each library was sequenced on an Illumina HiSeq2500 using Rapid mode with v2 chemistry as PE with 250-bp read lengths. This library was sequenced to a coverage of ~65 Gb (~30x the genome size). The 450-bp PCR-free shotgun library was sequenced on an Illumina HiSeq2500 using Rapid mode with v2 chemistry as PE with 250-bp read lengths. This library was sequenced to a coverage of ~130 Gb (~60x the genome size). In total, ~180x of sequencing data was produced for this project. All sequencing was conducted at the Roy J. Carver Biotechnology Center (Urbana, IL) at the University of Illinois.

**Genome assembly.** The W22 genome was assembled using DenovoMAGIC (2) and we provide a summary of the implementation of this approach for W22. Read pre-processing included removal of PCR duplicates, Illumina adaptor AGATCGGAAGAGCG and Nextera linkers (for MP libraries). For the 2x 250 bp PE libraries, overlapping reads were merged with a minimal required overlap of 10 bp to create stitched reads. Error correction of pre-processed reads was performed by scanning to detect reads with putative sequencing errors (containing a sub-sequence that does not reappear several times in other reads) and those reads were filtered out. The first step of de novo assembly consisted of building a de Bruijn graph (g) of 1.5×239 bp with the ‘stitched’ reads. The large k-mer size significantly reduced the complexity of the de Bruijn graphs, which is essential for high-quality assembly of very complex genomes. No bubble merge and no repeat masking filtering were used. To scaffold the contigs of the de Bruijn graph, non-repetitive contigs within the graph were identified and assembled into scaffolds based on mapping information of the ‘stitched’ reads. Scaffolding was completed using a directed graph containing scaffolds longer than 200 bp as nodes, and edges were based on the PE and MP links as vertices. Error-corrected consensus sequences were filtered out to generate unconnected sub-graphs that were ordered into scaffolds. PE reads were used to find reliable paths in the graph for additional repeat resolving. This was accomplished through searching the de Bruijn graph for a unique path connecting overlapping scaffolds to viral contigs which could then be connected to the scaffolds. These scaffolds were then further ordered and linked using the MP libraries, estimating gaps between the contigs according to the distance of MP links. Linking scaffolds with MP reads required confirmation of at least three filtered MPs or at least one filtered MP with supporting confirmation from two or more filter-failed MPs, where the NTPR was shorter than 50 bp. All scaffold edges were masked and links between non-repeative contigs mapping to the same scaffolds were united, generating a directed scaffold graph. Further ordering of scaffolding was achieved through alignment of the scaffolds to the B73 reference genome (RefGen_v2) and selection of the most probable genomic location. We did not use the B73 reference to create the W22 scaffolds. The scaffolds were then placed into the pseudomolecules to maximize linear synteny between the assembled and B73 genomes. To improve the accuracy of our scaffolds, we used an independent method to identify and correct misassemblies. We sequenced one HiSeq X10 lane (total of 120G of 150 × 2bp raw sequencing data) with one long-range barcoded DNA library (Chromium system by 10x genomics). Reads were mapped to the assembled scaffolds and clusters of reads with the same barcode mapped to adjacent contigs in the scaffolds were identified. Overall, we detected 22 million clusters and the estimated length of >2 million of those molecules is above 50 kb (the estimated average length of all molecules is 15 kb with s.d. = 23 kb). Next we scanned along each scaffold with a 20-kb-length window and counted: the number of distinct clusters that cover the whole window, the number of distinct clusters that cover the right edge of the window, and the number of distinct clusters that cover the entire window indicating a support for this 20 kb connection by several long molecules (identified by having distinct barcodes). On average, 255 long molecules (s.d. = 74) cover each edge of the window and 121 molecules (s.d. = 31) support an entire window (s.d. = 31). There are 34 potential chimeric scaffolds that were detected as windows with a statistically significant (using a hyper-geometric statistical test with a Bonferroni correction, P-value > 2.5 × 10⁻⁵) lower number of supporting overlapping long molecules. Twenty-three of these potential scaffolds were assembled into the GenomeScope scaffold validation stage (see below). These chimeric scaffolds were broken and reassembled in W22v2 using the X10 information.

**Bionano mapping and assembly validation.** High-molecular-weight genomic DNA was isolated from 0.5 g of W22 leaf tissue using the NripsPrep Plant Tissue DNA Isolation Kit (; the DNA was labeled at Nt.BspQI sites using the IrysPrep NRRS labeling kit (IrysGenome). Molecules were sequenced on the IrysChips at 100x coverage with an average molecule length of 220 kb. The data were then de novo assembled using opt/Argument human. The resulting W22 assembly contained 1,872 Bionano genome maps (equivalent to contigs that span a total of 2,171 Mb with an N50 of 6.7 Mb). After assembly, the W22 genome maps were aligned to an in silico BspQI-digested cmap of the NRGene sequence assembly. The final alignment parameter ‘Total Unique Aligned Lenv Ret Len’ was 0.983, indicating that 98.3% of the BioNano maps aligned uniquely with the NRGene assembly. As a means to directly compare the quality of the W22 assembly with the quality of the B73 assembly, we created a hybrid scaffold that incorporated both the sequence and Bionano data. The W22 hybrid assembly contained 97.4% of the Bionano contigs and 99.4% of the sequence assembly. These values are slightly better than equivalent values for B73 (95.1% and 98.4%, respectively), indicating that the ordering and orientation of contigs in the NRGene assembly are on par with the PacBio-based B73 reference assembly.

**Scaffold validation.** The 4.4 million high-resolution genetic anchors, which were developed from >14,000 maize inbred lines (5), were used to validate the scaffolds of the W22 assembly. The sequence anchors were aligned to an initial assembly of W22 using Bowtie2. The scaffolds with at least 20 mapped anchors were examined for the consistency of sequence assembly and genetic position. A total of 409 scaffolds with an overall length of 2.17 Gb were validated. We found that 23 scaffolds were assembly artifacts, which were then corrected in the W22v2 assembly.

**W22 genome annotation.** W22 gene annotation. Annotation of protein-coding genes was performed using MAKER pipeline software (19), with parameters and evidence similar to those recently used to annotate B73 (18, 22). Repeat masking (‘RepeatMasker Home Page’ 2017) was performed using exemplar transposon sequences available online at the maize TE database (‘Maize Database’ 2017). We excluded helitron and MULE elements to avoid false-positive masking from captured exon sequences in such elements. Gene expression evidence included PacBio iso-seq long reads sequenced from cdNA libraries of six tissues in B73 (n = 111,151) (18). In addition, we included the following transcriptome assemblies, each processed to exclude short transcripts (<0.3 bp) and redundancies based on annotation of CD-HIT (43): a portion of 94 transcripts extracted from publicly available RNA-seq reads (n = 508,233) (18); a transcriptome assembly of B73 seedlings (n = 112,963) (18); a transcriptome assembly of W22 tissues (n = 589,743). Cross-species evidence was supplied in the form of the following annotated protein files downloaded from Gramene release 46 (Gramene FTP) (47): Arabidopsis_thaliana.TAIR10.27.pep.all.fa; Brachypodium_dichotomum.v1.0.27.pep.all.fa; Oryza_sativa.IGRP.1.0.27.pep.all.fa; Setaria_italica.IGv02.7.0.27.pep.all.fa; Sorghum_bicolor.Sorbi.27.pep.all.fa. Alignment and downstream processing of sequence evidence to the repeat-masked W22 reference was performed within the MAKER pipeline using default parameters. For gene predictions, we used the pipeline incorporated AUGUSTUS+ (applied with the ‘maize5’ model and FGENESH+) applied with the ‘monocot’ model. Stable gene identifiers were assigned using the format Zm00004bXXXXXXX (where the Xs are a random 6-digit number), as specified by the maize community nomenclature standard. For gene annotations, we used detectMITE (55). For helitrons, we use the terminal 30 base pair at the 3′ end to assign the helitron. For TIR elements, we utilize TARGeT on MTEC TE models. LTRharvest was used to identify and correct misassemblies. We sequenced one HiSeq X10 lane (total of 120G of 150 × 2bp raw sequencing data) with one long-range barcoded DNA library (Chromium system by 10x genomics). Reads were mapped to the assembled scaffolds and clusters of reads with the same barcode mapped to adjacent contigs in the scaffolds were identified. Overall, we detected 22 million clusters and the estimated length of >2 million of those molecules is above 50 kb (the estimated average length of all molecules is 15 kb with s.d. = 23 kb). Next we scanned along each scaffold with a 20-kb-length window and counted: the number of distinct clusters that cover the whole window, the number of distinct clusters that cover the right edge of the window, and the number of distinct clusters that cover the entire window indicating a support for this 20 kb connection by several long molecules (identified by having distinct barcodes). On average, 255 long molecules (s.d. = 74) cover each edge of the window and 121 molecules (s.d. = 31) support an entire window (s.d. = 31). There are 34 potential chimeric scaffolds that were detected as windows with a statistically significant (using a hyper-geometric statistical test with a Bonferroni correction, P-value > 2.5 × 10⁻⁵) lower number of supporting overlapping long molecules. Twenty-three of these potential scaffolds were assembled into the GenomeScope scaffold validation stage (see below). These chimeric scaffolds were broken and reassembled in W22v2 using the X10 information.

**Characterization of variation in gene content and arrangement.** To compare syntenic ortholog retention between W22 and B73, we queried the unmasked W22 genome against the unmasked B73 version 4 from Ensembl release 36 and the outgroup
Sorghum bicolor v3.126 using CoGe's SynFind (lastz, 50-gene window, minimum of 5 syntelogs... those families with 10+ members were used to generate NATuRE GENETICS | www.nature.com/naturegenetics
Page 37

NATuRE genomes were each searched against themselves using COGE's SynFind tool. Locally duplicated genes were determined for Sorghum bicolor using the processing pipeline and placement quality criteria described previously. To compare alternative splicing in W22 and B73, we took the guided assembly; in addition, we also used de novo transcripts from Trinity. The two approaches, Stringtie 1.3.3 and Cufflinks 2.2.1, were used for the genome mapping back to the W22 genomes and called the alternative splicing loci in UniformMu lines were mapped in the B73v4 and W22v2 genomes using the master list of ortholog mappings using SynFind. Reciprocal SynFind searches were used where W22 was searched against B73 and vice versa. The reciprocal B73 to W22 gene model mappings that were found in both searches were filtered for high-confidence orthologs. Tool Farm duplications were a subset of the locally duplicated genes that had no intervening genes between them.

Analysis of gene expression. RNA seq alignments. RNA-seq data for W22 leaf (SRA: SRR1986767) were aligned to both the W22 and B73v4 reference genomes. Reads were trimmed using Cutadapt version 1.8.1 and mapped using tophat version 2.0.13 to both genomes. Stranded reads were assigned to gene models for each reference using HTSeq version 0.5.3. The counts per gene were normalized to reads per million mapped reads. To compare gene expression values for mappings to the two references, homologous genes were determined from predicted protein sequences for the first isoform of each gene using OrthoFinder version 1.1.8. Only genes with a 1-to-1 match between W22 and B73v4 were retained for comparison.

Alternative splicing analysis. W22 RNA-seq reads were trimmed using Cutadapt 1.7.1 to remove adapters and low quality sequences trimmed using Trimomatic 0.32. Processed reads were then aligned to the genome using STAR 2.5.2 and two pass alignment steps. Sixty-seven RNA-seq sample were aligned to STAR individuals, and splice junction from 67 samples were combined together and only splice junction >= 3 reads at least one sample kept. These remaining splice junctions were fed into STAR as guidance and the second alignment was run. Duplicate reads were detected and removed using Picard 2.10.3 MarkDuplicates. Two approaches, Stringtie 1.3.3 and Cufflinks 2.2.1, were used for the genome-guided assembly; in addition, we also used de novo transcripts from Trinity. The transcripts built from Cufflinks and Stringtie were filtered using the same approach previously. To compare alternative splicing in W22 and B73, we took the transcripts from the Stringtie annotation that includes alternative splicing isoforms, and mapped back to the W22v2 genomes and called the alternative splicing loci in B73 relative to the W22v2 genome.

Mapping of Mu andDs insertion sites. Germinal Mu insertions from 13,444 UniformMu lines were mapped in the B73v4 and W22v2 genomes using the Illumina-based MuSeq protocol as described in an earlier study. Insertion sites that were assigned unique locations in the W22 genome were analyzed to determine the presence and locations of corresponding sites in the B73v4 genome. Correspondence of insertion sites in W22 and B73 was based on shared MuSeq reads. Insertion sites in W22 identified by MuSeq reads that failed to align to B73 were scored as W22-specific. Flanking Ds (ID) sequences, which derive from sequenced clones, were placed on each genome assembly using the processing pipeline and placement quality criteria described previously, with minor modifications. For each genome, the pipeline was first applied to the unmethylated genome and unmasked ID sequences, and then the remaining, unplaced ID sequences were repeat masked and run through the pipeline again with the unmethylated genome. Multiple ID insertions that were collapsed into a single placement when at least one ID clone included a sequence directly adjacent to the Ds insertion site. Repeat masking used Repeatmasker as described previously.

DNA methylation and chromatin accessibility. Whole-genome bisulfite sequencing and analysis. Genomic DNA (1 µg) was sheared to a size of 200–300 bp. These DNA fragments were then used to construct a whole-genome bisulfite sequencing library using the KAPA library preparation kit (KK6232). Briefly, the DNA fragments were treated with sodium bisulfite and then purified using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), using the manufacturer’s instructions. Indexed libraries were pooled and sequenced on four Illumina HiSeq 2500 lanes with PE 50-cycle sequencing. Data are available in NCBI SRA (accession SRP118121).

Metaplots for chromatin surrounding genes or TE insertion sites. For DNA methylation, the context-specific levels of DNA methylation for each 100-bp bin across the genome were utilized. For chromatin accessibility, the normalized (reads per million) values for the heavy and light digest were calculated for each 100-bp bin across the genome were utilized. For chromatin accessibility, the normalized (reads per million) values for the heavy and light digest were calculated for each 100-bp bin across the genome were utilized. For chromatin accessibility, the normalized (reads per million) values for the heavy and light digest were calculated for each 100-bp bin across the genome were utilized.
the figure; see Supplementary Table 8 for details on these families), the average level of DNA methylation or chromatin accessibility was calculated for each 100-bp window with respect to the TE edge. Mu and Ds data were then added for each DNA methylation context and MNase. The heatmap was then generated using a Euclidean clustering system focusing on the CG and CHG methylation levels. Once the family order was determined, a separate scale was used to generate each context of methylation and MNase to view the trend within each (CG/CHG = 0–1, CHH = 0–0.15, MNase = 0–4).

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** The W22v2 genome sequence is available at NCBI under the accession GCA_001644905.2. The W22 annotations are available at MaizeGDB.org. There are also RNA-seq (SRR1986376), whole-genome bisulfite (SRX3136383) and MNase (SRP118121) data sets available at NCBI SRA.

**References**

Experimental design

1. Sample size
   Describe how sample size was determined. For the primary experiments the sample size was 1 (for the genome assembly and annotation). For other experiments the sample size was the population of lines in which novel TE insertions had been characterized.

2. Data exclusions
   Describe any data exclusions. No data was excluded

3. Replication
   Describe whether the experimental findings were reliably reproduced. Most experiments were not replicated as they report the results of genome assembly and annotation. Chromatin data was assessed through biological replication and comparisons of profiles at multiple sites.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups. Not relevant to this study

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis. Blinding was not relevant for this study

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

<table>
<thead>
<tr>
<th>n/a</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>☑✓</td>
<td>The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)</td>
</tr>
<tr>
<td>☑</td>
<td>A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.</td>
</tr>
<tr>
<td>☑</td>
<td>A statement indicating how many times each experiment was replicated</td>
</tr>
<tr>
<td>☑✓</td>
<td>The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)</td>
</tr>
<tr>
<td>☑✓</td>
<td>A description of any assumptions or corrections, such as an adjustment for multiple comparisons</td>
</tr>
<tr>
<td>☑✓</td>
<td>The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted</td>
</tr>
<tr>
<td>☑</td>
<td>A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)</td>
</tr>
<tr>
<td>☑✓</td>
<td>Clearly defined error bars</td>
</tr>
</tbody>
</table>

See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study. The DeNovoMagic2 is proprietary software developed by NRGene. A summary of the method is detailed in Online Methods. Other software used in this study is publicly available and described.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Seed was sourced from Dr. Hugo Dooner who maintained the stocks and deposited materials at the USDA GRIN under accession PI674445. This stock is freely available at: https://training.ars-grin.gov/gringlobal/AccessionDetail.aspx?id=1925431

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

b. Describe the method of cell line authentication used.

c. Report whether the cell lines were tested for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Research did not involve human subjects
Figure 1. Scheme of biosynthetic pathways from Indole-3-glycerolphosphate to DIBOA-Glc, Tryptophan and free indole. Confirmation of IGPS1, IGPS2, and IGPS3 enzymatic activity is described in this manuscript. BX = benzoxazinoneless, TSA = tryptophan synthase alpha subunit, TSB = tryptophan synthase beta subunit, IGL = indole-3-glycerolphosphate lyase, DIBOA-Glc = 2,4-dihydroxy-1,4-benzoxazin-3-one glucoside, ER = endoplasmic reticulum.
Figure 2. Identification of maize indole-3-glycerol-phosphate synthases

A Dendrogram analysis of indole-3-glycerol-phosphate synthases (IGPSs, orange) and indole-3-glycerol-phosphate lyases (green), and tryptophan synthases (blue) from maize, Arabidopsis, and E.coli. Maize BX2 and BX3 were used as an outgroup. B IGPSs enzyme activity was verified by complementation of an tryptophan auxotrophic phenotype.
Figure 3 Transcript abundance if *IGPS1*, *IGPS2* and *IGPS3* to stress responses in *W22* leaves.

Transcript levels of the IGPSs genes were determined after (A) *Spodoptera Exigua* feeding (mean +/- s.e. of n = 5), (B) methyl jasmonate induction (mean +/- s.e. of N = 3), and (C) aphid infestation (mean +/- s.e. of N = 8). Values are fold change to uninfested controls (n = 5). *P < 0.05, ***P < 0.001 by Student’s t-test.*
Figure 4 Effects of igps1 knock-out mutant on gene expression, metabolites and caterpillar growth.

Transcript levels of IGPS synthase genes in a (A) segregating and (B) homozygous igps1 mutant line (mean +/- s.e of N = 6). Comparison of the five most abundant (B) benzoxazinoids (mean +/- s.e of N = 4-8) and (D) tryptophan (mean +/- s.e of N = 4-8) in a segregating igps1 mutant population. (E) Performance of Spodoptera exigua on igps mutant vs. wildtype W22. *, P < 0.05, **, P < 0.01; ***, P < 0.001 by Student’s t test. ANOVA, Fischer’s LSD 0.05
Figure 5 Effects of *igps2* mutant line on gene expression and metabolite changes. (A) Transcript levels of IGP synthase genes in homozygous *igps2* mutant line vs W22 (mean +/- s.e. of N = 4). (B) Comparison of tryptophan (mean +/- s.e. of N = 4) and (C) the five most abundant benzoxazinoids (mean +/- s.e. of N = 4). *P < 0.05, **P < 0.01; ***P < 0.001 by Student’s t test.
Figure 6 Cellular localization of IGP synthases, BX1, BX2, TSA and IGL.
(A) Images of maize mesophyll protoplasts transiently expressing 35S-TSA-YFP, 35S-IGPS2-YFP, 35D-IGL-YFP, 35S-IGPS1-YFP, 35S-BX1-YFP, 35S-IGPS3-YFP showing YFP fluorescence in chloroplasts, chlorophyll auto-fluorescence and merged images. (B) Co-localization of 35S-BX2-mcherry with ER-CFP.
Figure 7. Interactions of IGPSs with BX1, TSA, and IGL by BiFC assays in *N. bentamiana* cells.
The overlap of YFP (yellow) and the chloroplast auto-fluorescence (red) images is shown in the column labeled merged. IGPSs were fused with C-terminus of YFP (c-YFP) and each construct was co-transformed into *N. bentamiana* leaves respectively, with BX1, TSA, and IGL fused with N-terminus of YFP (n-YFP).
Notable results: regulation of photomorphogenesis by benzoxazinoids

One of the initial hypotheses to be tested in this research was that benzoxazinoids directly interfere with auxin signaling. Neither the Chamovitz nor the Jander lab was able to verify this effect in maize or Arabidopsis. Nevertheless, we observed a benzoxazinoid effect on blue light-induced curvature of maize seedlings (Figures 1, 2), which is known to be regulated by auxin. Our current hypothesis, supported by several independent experiments, is that the differential seedling curvature is caused by benzoxazinoid-regulated differential regulation of the expression of \( NPH3 \), a gene that encodes a component of the plant blue light receptor, in plants with and without benzoxazinoids.

The hypothesis that DIMBOA-regulated gene expression mediates photomorphogenic responses in maize needs to be verified using analysis of maize \( NPH3 \) mutant lines. Once this has been accomplished, we will publish our results.

Figure 1. Setup for detecting blue light-induced hypocotyl curvature. Maize seedlings illuminated from the side with blue light initiate auxin-dependent hypocotyl curvature.

Figure 2. Maize without benzoxazinoids exhibits delayed curvature. Maize seedlings were illuminated from one side with blue light and the rate of curvature was measured over two hours.
Figure 3. NPH3 (nonphotomorphic hypocotyl 3) expression is reduced in maize bx1 and bx2 mutant plants relative to wildtype inbred line W22. NPH3, which is required for normal blue light responses, is expressed at a lower level in maize lines that lack benzoxazinoids.
Notable results: regulation of secondary metabolism by benzoxazinoids

Both transcript profiling (Figure 1) and metabolite profiling (Figure 2) show broad differences between wildtype maize and mutant plants ($bx1$ and $bx2$) that lack benzoxazinoids. Further work was conducted to identify specific metabolites that are differentially regulated by benzoxazinoids but are not metabolically related. Analysis of the transcriptomic and metabolomic data identified both up- and down-regulation that was dependent on the presence of benzoxazinoids.

Examples of three compounds that are more abundant in benzoxazinoid-containing plants are shown in Figure 3. Further experiments identified catechol, along with its acetylated and glucosylated derivatives as being up-regulated by benzoxazinoids. These compounds are likely to be defense-related, but are made via a biosynthetic pathway that is distinct from that of benzoxazinoids. Further research with other funding is focused on identifying the mechanisms of benzoxazinoid-regulated changes in plant metabolism.

**Figure 1. DIMBOA influences maize gene expression.** Principal component analysis shows broad differences in gene expression changes between wildtype W22 and $bx1$ mutant maize, with and without insect feeding for 0, 2, 8, and 24 hours.
Figure 2. PCA shows a different metabolite content in bx1 and bx2 mutants relative to wildtype W22 and a1 mutant maize. Non-targeted HPLC-MS metabolomic analysis was conducted, with and without insect feeding.

Figure 3. Compounds that are differentially regulated by the absence of benzoazinoids. Examples of three compounds that are more abundant in benzoazinoid-containing a1 and W22 plants than in otherwise isogenic benzoazinoid-mutant bx1 and bx2 plants. 109.0280363, 313.0920342 and 271.0792271 are masses of compounds that were determined by quadrupole-orbitrap mass spectrometry. Data are means +/- standard error of N = 5.