

Document downloaded from:

<http://hdl.handle.net/10251/65386>

This paper must be cited as:

Sarrion-Perdigones, A.; Vázquez Vilar, M.; Palací Bataller, J.; Castelijns, B.; Forment Millet, JJ.; Ziarsolo Areitioaurtena, P.; Blanca Postigo, JM.... (2013). GoldenBraid 2.0: a comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiology*. 162(3):1618-1631. doi:10.1104/pp.113.217661.



The final publication is available at

<https://dx.doi.org/10.1104/pp.113.217661>

Copyright American Society of Plant Biologists

Additional Information

**Running Head**

GoldenBraid2.0

**Corresponding Author:**

Diego Orzaez

Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia. Avda Tarongers SN, 46022 Valencia, Spain

Email: [dorzaez@ibmcp.upv.es](mailto:dorzaez@ibmcp.upv.es)

Tel: +34-963879933

**Research Area:**

Breakthrough Technologies

**Title**

GOLDENBRAID 2.0: A comprehensive DNA assembly framework for Plant Synthetic Biology

**Authors**

Alejandro Sarrion-Perdigones<sup>\*1</sup>, Marta Vazquez-Vilar<sup>\*1</sup>, Jorge Palací<sup>1</sup>, Bas Castelijns<sup>1</sup>, Javier Forment<sup>1</sup>, Peio Ziarsolo<sup>2</sup>, José Blanca<sup>2</sup>, Antonio Granell<sup>1</sup> and Diego Orzaez<sup>1</sup>

**Address:**

(1) Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia. Avda Tarongers SN, 46022 Valencia, Spain

(2) Centro de Conservación y Mejora de la Agrodiversidad Valenciana. Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain.

**One-Sentence summary**

GoldenBraid2.0 is a comprehensive technological framework that facilitates the construction of increasingly complex multigene structures at the DNA level while enabling the exchange of genetic building blocks among plant bioengineers

**Corresponding Author:**

Diego Orzaez

Instituto de Biología Molecular y Celular de Plantas (IBMCP), Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia. Avda Tarongers SN, 46022 Valencia, Spain

Email: [dorzaez@ibmcp.upv.es](mailto:dorzaez@ibmcp.upv.es)

Tel: +34-963879933

## **ABSTRACT**

Plant Synthetic Biology aims to apply engineering principles to plant genetic design. One strategic requirement of Plant Synthetic Biology is the adoption of common standardized technologies that facilitate the construction of increasingly complex multigene structures at the DNA level while enabling the exchange of genetic building blocks among plant bioengineers. Here we describe GoldenBraid2.0 (GB2.0), a comprehensive technological framework that aims to foster the exchange of standard DNA parts for Plant Synthetic Biology. GB2.0 relies on the use of TypeIIIS restriction enzymes for DNA assembly and proposes a modular cloning schema with positional notation that resembles the grammar of natural languages. Apart from providing an optimized cloning strategy that generates fully exchangeable genetic elements for multigene engineering, the GB2.0 toolkit offers an ever-growing open collection of DNA parts, including a group of functionally-tested, pre-made genetic modules to build frequently-used modules like constitutive and inducible expression cassettes, endogenous gene silencing and protein-protein interaction tools, etc. Use of the GB2.0 framework is facilitated by a number of web resources which include a publicly available database, tutorials and a software package that provides *in silico* simulations and lab protocols for GB2.0 part domestication and multigene engineering. In short, GB2.0 provides a framework to exchange both information and physical DNA elements among bioengineers to help implement Plant Synthetic Biology projects.

## INTRODUCTION

Synthetic Biology is producing a paradigm shift in Biotechnology based on the introduction of engineering principles in the design of new organisms by genetic modification (Check, 2005; Haseloff and Ajioka, 2009). Whereas Synthetic Biology has rapidly permeated microbial biotechnology, the engineering of multi-celled organisms following Synthetic Biology principles is now emerging, and is mainly driven by the so-called top-down approaches where newly engineered genetic circuits are embedded into naturally-existing organisms used as a “chassis”. The plant chassis offers an extraordinarily fertile ground for Synthetic Biology-like engineering. However, technology still faces the huge challenge of performing engineering-driven genetic designs. One of the main technological challenges of Plant Synthetic Biology requires the construction and transfer of multigene structures to the plant genome. This is putting pressure on developing a DNA assembly and transformation technologies adapted to plants. One main trend is the use of modular cloning, an engineering-inspired strategy consisting in the fabrication of new devices by combining prefabricated standard modules. In a modular strategy, pre-defined categories, the so-called “parts”, are assembled together following a number of rules known as the “assembly standard”. Modular DNA building has been enthusiastically adopted by microbial Synthetic Biologists because it offers a number of advantages such as speed, versatility, lab autonomy, combinatorial potential, and often lower cost (Ellis et al., 2011). Modular methods acquire full potential when parts are easily interchangeable, and when one or a few assembly standards are shared by many manufacturers.

A number of features define the value of a modular cloning method. Speed and efficiency are important characteristics, as are also its simplicity and the ability to produce scar-less or scar-benign assemblies. Moreover, any cloning strategy for Synthetic Biology should enable endless reusability; that is, it should ensure that new composite parts themselves can take part in new assemblies, therefore allowing unlimited growth. Several modular cloning strategies have been proposed in the literature, and each presents advantages and shortcomings. For instance, the original BioBrick standard widely used in microbial Synthetic Biology scores a maximum for simplicity because a single rule governs all the assemblies (a property known as idempotency). However, it is not scar-benign and is only relatively efficient (Knight, 2003). LIC (Aslanidis and de Jong, 1990), USER's (Geu-Flores et al., 2007), and specially Gibson Assembly (Gibson et al., 2009), are highly efficient DNA assembly methods, although they are neither strictly modular nor widely adopted by plant biotechnologists. In sharp contrast, Gateway Cloning (Hartley, Temple et al. 2000) is of widespread use in plant laboratories (Karimi et al., 2002; Karimi et al., 2007; Estornell et al., 2009). Recently, MultiRound Gateway technologies opened Gateway capabilities to the sequential delivery of multiple transgenes by multiple rounds of recombination reactions (Chen et al., 2006; Vemanna et al., 2013). In general, Gateway-based technologies are highly efficient. Unfortunately, they are not always scare-benign as they leave 21 bp scars between building blocks. Other technologies involving rare cutters or homing endonucleases-

based strategies have also been developed and adapted to plant transformation (Lin et al., 2003; Dafny-Yelin and Tzfira, 2007; Fujisawa et al., 2009), including combinations of homing endonucleases and engineered zinc finger nucleases (Zeevi et al., 2012), and iterative *in vivo* assembly rounds of Cre recombinase and phage1 site-specific recombination (Chen et al., 2010). Many of these techniques can serve as efficient assembly methods for multigene engineering. Nonetheless, a pre-requisite to become a standard for Plant Synthetic Biology is the development of a set of rules and tools based on those technologies which can be shared by as many labs as possible.

Recently, a very powerful DNA assembly method named Golden Gate was described (Engler et al., 2008; Engler et al., 2009). Golden Gate uses Type IIS restriction enzymes to generate four-nucleotide sticky ends flanking each DNA piece, which can be subsequently joined together efficiently by T4 ligase. The assembly reaction is multipartite and is performed in a single tube reaction to yield highly efficient scar-less or scar-benign assemblies. This is because Type IIS recognition sites are eliminated upon ligation, leaving only four nucleotides seams, which can be user-defined. These features make the Golden Gate technology an excellent candidate to set up a standardized Modular Cloning system. However, as originally conceived, Golden Gate is not a reusable system and cannot, therefore, be used efficiently for multigene engineering.

Most recently, two strategies were described to enable the reusability of the Golden Gate cloning scheme: MoClo (Weber et al., 2011) and GoldenBraid (Sarrion-Perdigones et al., 2011). Both methods use the multipartite Golden Gate property to build transcriptional units (TUs) starting from basic standard building blocks, and both create specially-designed destination vectors to enable Golden Gate-built TUs to be assembled among them. Whereas the GoldenBraid minimalist cloning strategy allows multigene growth by enabling binary assemblies between TUs, the MoClo destination vectors offer the interesting possibility of performing multipartite assemblies at the TU level, be it at the cost of the higher complexity of its vectors toolkit.

The Golden Gate-based strategies MoClo and GoldenBraid are ideal to serve as modular assembly systems in Plant Synthetic Biology as they are efficient, reusable and scar-benign. To realize their full potential, it is very important to: (i) advance in adopting common standards; so building blocks can be shared by as many users as possible; (ii) further optimize the design of cloning strategies to improve speed and efficiency; (iii) improve users' experience by generating new hardware (building blocks and modules) and software (databases and assembly programs) tools which simplify and facilitate the engineering process.

To facilitate the implementation of Plant Synthetic Biology approaches, we present GoldenBraid 2.0 (GB2.0), a new version of the GoldenBraid cloning strategy. In this new version, we defined, in concert with MoClo developers, a common assembly standard by establishing arbitrary, yet scar-benign, assembly seams within a TU which facilitates part exchangeability. In addition, we

optimized the versatility of the GB strategy by enhancing its minimalist design, creating a universal part entry vector and simplifying the cloning setup. Finally, we generated a collection of pre-made genetic modules and new software tools for the purpose of facilitating the building of frequently used genetic structures. In short, we present a new grammar for Plant Synthetic Biology and we introduce a comprehensive toolkit to facilitate the use of GB2.0 in composing new genetic designs.

## RESULTS

### The GB2.0 cloning strategy

To describe the GB2.0 assembly strategy, we follow an analogy with a natural language because we believe this comparison closely describes the GB2.0 cloning strategy structure and facilitates its understanding. This is because the hierarchical manner in which the different building blocks in GB2.0 are combined to form a multigenic structure become analogous to the way grammar elements (morphemes, words, phrases and sentences) are combined hierarchically to create a composition. Figure 1A provides an equivalence table between the elements of English grammar and the elements of the GB2.0 system.

#### *GBparts: words and phrases. Definition of the GB2.0 grammar*

The first task in upgrading GoldenBraid was to define the minimal standard building blocks in GB2.0, the so-called GBparts, which can be considered the “words” of the GB grammar. GBparts are fragments of DNA flanked by four nucleotide overhangs. They are stored as inserts within a specially designed entry vector (pUPD), from where they are released by cleavage with Bsal or BtgZI restriction enzymes to generate the corresponding flanking overhangs. GBparts are classified into different classes or categories according to their specific function. Each GB class is defined by its flanking four nucleotides which will overhang upon enzyme digestion and will determine its position within the TU. We defined eleven standard classes (Figure 2A), which correspond to the basic functional categories in a typical TU. The first three categories (01, 02 and 03) were orderly set in the 5´ non-transcribed region, and correspond to operators and promoter regions. Next, we defined seven categories in the transcribed region: one corresponded to the 5´ UTR (11); one related to the 3´ UTR (17); four were reserved to the coding region (13-14-15-16); an additional class was set as a buffer zone to facilitate, among other designs, the construction of non coding TUs intended for gene silencing. Lastly, we set a final class (21) for standard 3´ un-transcribed GBparts.

Besides the basic classes, GB2.0 also employs “superclasses”. For practical purposes, it is convenient to group several contiguous basic GBparts which, together, perform a defined function (e.g., a complete promoter or a full coding region) in a single DNA element (a GB super-part, abbreviated to GBSpert) instead of splitting it into its basic standard parts. This is analog to an English phrase, which comprises a group of words that functions as a single unit within the hierarchical structure of the sentence syntax (e.g., a subject or a direct complement). As with GBparts, GBSperts are ultimately DNA fragments stored within the pUPD vector. Upon digestion with Bsal or BtgZI, the whole phrase is released as a solid indivisible unit flanked by four nucleotides barcodes. In practice, GBSperts are very convenient as they reduce the number of elements that need to be assembled to produce a TU and, therefore, they enhance efficiency. Frequently used superclasses are depicted in Figure 2B and C. For example, the promoter regions normally employed in traditional cloning correspond to the superclass (01-12).

GBparts and GBSparts are components of the GB collection, and their sequence information is stored in the GBdatabase.

#### *GBpart domestication: creating words and phrases*

The process of adapting a DNA building block (GBparts or GBSparts) to the GB grammar is referred to as domestication. GB domestication usually involves the PCR amplification of the target DNA (word or phrase) using GB-adapted primers (see Figure 3 for details), and the subsequent cloning of the resulting PCR fragment into the pUPD vector using a BsmBI restriction-ligation reaction. Occasionally, domestication may involve the removal of internal BsaI, BsmBI or BtgZI restriction sites. In order to facilitate an eventual automation of the cloning process, the GB2.0 system includes a standard procedure for internal site removal. This procedure, described in detail in Supplementary Figure 1, involves the amplification of the target DNA in separated fragments (named GBpatches) using GB-adapted primers, which incorporate single mismatches to disrupt the enzyme target sites. Once amplified, GBpatches are re-assembled together in a single-tube BsmBI restriction-ligation reaction into pUPD to yield a domesticated GBpart or GBSpart.

#### *The GB2.0 destination plasmids kit*

GoldenBraid destination vectors (pDGBs) are binary vectors that function as recipients of new assemblies. Each pDGB contains a GBCassette (the selection lacZ gene flanked by two restriction/recognition sites corresponding to two different type IIS enzymes; see Figure 4A). In addition, GB2.0 plasmids include a watermark (i.e., a distinctive restriction site flanking the GBCassette) to help plasmid identification. Detailed information about the sequence of the different GBCassettes is also provided in Figure 4A. The special orientation and arrangement of the restriction enzymes defines two levels of pDGBs; the  $\alpha$ -level and  $\Omega$ -level plasmids; which are used for the BsaI and BsmBI-GB reactions, respectively. Plasmids also differ in the resistance marker that is associated with each level (kanamycin for level  $\alpha$  and spectinomycin for level  $\Omega$ , allowing counter-selection). To ensure an endless cloning design, a minimum set of four pDGBs is required (pDGB $\Omega$ 1, pDGB $\Omega$ 2, pDGB $\alpha$ 1 and pDGB $\alpha$ 2). Additionally, this set can be expanded to eight plasmids to enable assemblies in different orientations (pDGB $\Omega$ 1R, pDGB $\Omega$ 2R, and pDGB $\alpha$ 1R and pDGB $\alpha$ 2R). For GB2.0, we constructed two complete sets of pDGBs, one based on the pGreen-II backbone and another set based on the pCAMBIA backbone. The sequence information of all 16 pDGBs in GB2.0 is uploaded in the GBdatabase.

#### *The composing strategy: from single words to full compositions*

The GB2.0 cloning strategy comprises two types of assemblies (see the GB2.0 chart in Figure 1B): multipartite assemblies and binary assemblies. Multipartite assemblies are performed to create single TUs. The different GBparts and GBSparts required to produce a well-constructed TU are mixed together in a single tube in the presence of a pDGB, the corresponding type IIS restriction enzyme/s, and the T4 ligase, and they are incubated in cyclic restriction-ligation

reactions. If all the elements are correctly set in the reaction, they orderly assemble within the destination vector and generate a so-called expression vector, which harbors the assembled composite part. Our pDGBs are binary vectors, therefore the resulting expression clone is ready to be used directly for *Agrobacterium*-mediated plant transformation.

After building a new TU using a multipartite assembly, the resulting new expression clone can be binarily combined with another expression clone to produce increasingly complex multigene structures analogously to how sentences are combined to create a written composition. The solution provided by GB cloning relies on the special design of GB destination vectors, which introduces a double loop (braid) into the cloning strategy. A composite part (a TU or a group of TUs) cloned in a given entry vector can be combined only with a second composite part cloned in the complementary entry vectors at the same level. This is done in the presence of a destination vector of the opposite level and generates a new expression vector at the opposite level. A formal notation describing the rules for multipartite and binary assemblies is shown in Figure 4B and C.

By choosing appropriate combinations of expression and destination vectors, it is possible to create increasingly complex structures, and the only limits are the capacity of the vector backbone or the biological restrictions imposed by bacteria. Moreover, all the new composite parts are fully reusable (they can be used directly for part transformation or can be employed in new assemblies) and exchangeable (can be combined with the GB modules that are produced separately in different labs by following the same assembly rules).

#### *Innovative features in the GB2.0 cloning strategy.*

Besides a proposal for a grammar, GB2.0 introduces a number of new elements that modify the original GoldenBraid cloning design to make it simpler and more versatile. Many of the new GB2.0 features rely on the design of the plasmid that harbors GBparts and GBSparts, the Universal Domesticator (pUPD). The pUPD cassette is designed to serve as a polyvalent entry vector for all the different GBparts and GBSparts, regardless of their category. This is because the four nucleotide barcodes are incorporated into the GBpart by PCR instead of being imprinted in the plasmid itself. Such a universal plasmid enables us to establish a single standard protocol for all the domestication parts based solely on its sequence information and category specification.

Another innovative feature of pUPD is the incorporation of both BtgZI and BsaI sites flanking the GB cassette. The enzyme target sites are arranged in such way that both BsaI and BtgZI digestions release exactly the same piece of DNA which contains the same four nucleotide overhangs, regardless of the enzyme used. This opens up the possibility of GBparts being assembled into the  $\alpha$  and  $\Omega$  level vectors indistinctly by using either BsaI-reactions or BtgZI/BsmBI-reactions, respectively. To enable this option, the GB cassettes in the pDGBs have also been redesigned and simplified. In the previous version, the sequences of the restriction sites for BsaI (named A, B and C) differed from the restriction sites for BsmBI (named

1, 2 and 3). In GB2.0, we made  $A \equiv 1$ ,  $B \equiv 2$  and  $C \equiv 3$  (see Figure 4A for details). In this way, and by making full use of the dual Bsal/BtgZI release from pUPD, any pDGB can be used as a recipient of a multipartite assembly which, therefore, makes entry in the GB loop fully symmetric. Thus, Bsal reactions are performed to build TUs in  $\alpha$ -vectors, and BsmBI/BtgZI reactions (BsmBI to open pDGB and BtgZI to release the GBpart) are performed to build TUs in  $\Omega$ -pDGBs. Furthermore by choosing any of the reverse pDGB plasmids as recipients, TU orientation can be inverted. This opens up the possibility of creating new binary assemblies in all the possible relative orientations.

The pUPD design provides yet another interesting new feature to GB2.0 as it enables the use of a non standard assembly level operating below the standard GBpart level (referred to as the GBpatch level). This feature can be most convenient for a number of applications, including the generation of seamless junctures, introducing combinatorial arrangements into protein engineering, or for promoter tinkering using non standard positions. The process is similar to the above-described domestication procedure. An example of the use of the GBpatch level for combinatorial antibody engineering is depicted in Supplementary Figure 2.

### **Frequently used structures.**

There is a limited number of structural types for the majority of synthetic transcriptional units and genetic modules. For instance, many protein-encoding TUs can be constitutively expressed, whereas others are regulated by 5' (or 3') operators. The resulting proteins can be preceded by a signal peptide, or may contain C-terminal and N-terminal fusions. Besides, noncoding TUs can be used for silencing purposes. To cope with this functional diversity while simplifying the users' toolbox, we defined a group of "Frequently Used Structures", for which specific pre-arranged GBparts and GBSparts were developed (depicted in Figure 2B and C). We now go on to describe some of the Frequently Used Structures that are currently included in the GB system and their associated tools.

#### *Basic expression cassettes for multigene engineering.*

Multigene engineering may require the use of different regulatory regions to avoid the silencing associated with the repeated use of a DNA sequence in the same construct. To meet this requirement, we incorporated several regulatory 5' and 3' regions into the GB2.0 collection. Most 5' regulatory regions are (01-12) GBSparts comprising a promoter and 5'-UTR, whereas 3' regulatory regions are (17-21) GBSparts comprising 3'-UTR and terminator regions. According to this basic set up, full (13-16) ORFs can be easily incorporated into tripartite reactions to obtain constitutively expressed TUs. In order to undertake Synthetic Biology projects, it is very important to have a range of regulatory regions available, and that the expression strength provided by each promoter/terminator combination is properly characterized so that the multigene expression can be adjusted accordingly. As a first approach toward the characterization of a set of basic expression cassettes, we finely characterized the

relative promoter/terminator strength of a number of cassettes using the Renilla/Luciferase system in transiently-transformed *N. benthamiana* leaves. The characterization of (01-12) and (17-21) regions as individual entities is a relatively straightforward procedure using GB2.0 cloning. However as the collection grows, the individual characterization of all the possible combinations becomes an intractable task. We therefore decided to investigate to what extent the transcriptional strength provided by each “promoter/terminator” (i.e., 01-12\_17-21) combination can be inferred from the separated contribution of each region. For this purpose, all the (01-12) promoter regions in the collection were tested by the Renilla/Luciferase system in combination with a common (17-21) terminator region (TNos). In parallel, all the (17-21) terminator regions in the collection were tested in combination with a common (01-12) promoter region (PNos). The “Experimental Transcriptional Activity” (ETA) of each region was calculated as being relative to the Renilla/Luciferase values of a (01-12\_17-21) reference combination (PNos\_TNos), which was arbitrarily set as 1 (see Figure 5A for the construct details). The ETA(01-12) values ranged between  $0.47 \pm 0.01$  and  $15.03 \pm 1.44$  relative luminescence units, whereas the ETA(17-21) values ranged between  $0.77 \pm 0.18$  and  $2.61 \pm 0.54$  (Figure 5B and 5C). Using these data, “Theoretical Transcriptional Activity” (TTA) was calculated for each cassette combination (Figure 5D) as the product of the individual ETA of the two regulatory regions. Finally, the Renilla/Luciferase ratio of a number of cassette combinations (covering 65% of total possibilities) were also tested experimentally. As we can see in Figure 5E and Supplementary Figure 3, there is a good agreement between the theoretical and experimental activity values. Of the 34 experimental combinations assayed in the evaluation test, 31 showed deviation in relation to the theoretical values below 2-fold ( $\pm 0.3$  in logarithmic values; for detailed information, see Supplementary Information 3C).

#### *Regulated expression cassettes*

The GB grammar contains several standard positions for the insertion of regulatory regions. In the 5' un-transcribed region, we defined three standard GBparts to allow combinatorial promoter tinkering and to facilitate the insertion of synthetic operators. As a functionality proof, we assembled and tested the pre-made cassettes for heat shock and the dexametasone-regulated expression; the latter is based on the “operated promoter A” scheme shown in Figure 2B. The Renilla/Luciferase/p19 reporter cassettes constructed with promoters pHSP70 and pHSP18.2 showed clear induction after incubation at 37°C (Supplementary Figure 4). The potential of the GB modular assembly was further demonstrated with the construction of two regulated systems based on the fusion of the glucocorticoid receptor (GR) with the DNA binding domains (BD) of LacI or Gal4 and the activation domain (AD) of Gal4. In this transactivation example, up to 15 pre-made modules comprising coding and noncoding regulatory regions were efficiently assembled *de novo* to produce two operated luciferase TUs which clearly responded to the presence of dexametasone (Supplementary Figure 5).

#### *Protein-protein interaction tools*

Reporter fusion partners are powerful analytical tools utilized in the study of protein-protein interactions. However, the use of unlinked co-transformation for the delivery of the interaction partners often compromises the extraction of reliable qualitative data, based on the poorly supported assumption that co-transformation efficiency in each cell is the same for all fusion partners. We reasoned that the linked co-transformation of fusion partners can help improve the sensitivity and accuracy of the protein-protein interaction analysis. By bearing this use in mind, we designed pre-made modules for the Bifluorescent Complementation assays (BiFC). For this purpose, BiFC adaptors with a (01-12) structure were constructed containing the full 35S promoter and the corresponding YFP or luciferase fusion partners. Based on this set-up, baits and preys with a canonical (13-16) structure can be easily assembled in multipartite reactions to form the required fusion proteins. The prearranged BiFC tools were functionally tested using transcription factors Akin10/Akin $\beta$ 2 as positive interaction partners, and an spermidine synthase (SPDS) as a negative partner (Belda-Palazon et al., 2012). As observed in Supplementary Figure 6, the number of cells showing positive interactions with the GB-assisted linked co-transformation set up outnumbers those of the unlinked co-transformation approach.

### *Silencing tools*

The negative regulation of endogenous genes often proves an engineering requirement. For this reason, special Frequently Used Structures were defined for three RNA silencing strategies: trans-acting small interfering RNAs (tasiRNA); artificial micro RNA (amiRNA); hairpin RNA (hpRNA) (Supplementary Figure 7A). Details of all the elements used in the RNAi designs are provided in Supplementary Table 1.

For the generation of tasiRNA constructs, special (01-11) GBSparts containing the mir173 trigger sequence are required. A CaM35S-based GBSpert for the constitutive tasiRNA expression is currently available in the GB collection. A regulated or tissue-specific tasiRNA expression can be designed using the GBpatch special feature of GB2.0. For the functional characterization of the tasiRNA structure, a 410-bp fragment of *A.thaliana* phytoene desaturase (PDS) (Felippes et al., 2012) was incorporated as a (12-16) GBSpert and was transformed into *A. thaliana* to yield approximately 0.1% seedlings with the albino phenotype (Supplementary Figure 7C). TasiRNA constructs require the co-expression of miR173 for effective silencing in plant species other than *Arabidopsis* (Felippes et al., 2012). To extend the species range of the tasiRNA tool, a new TU with a constitutively expressed miR173 was constructed and incorporated into the collection. The functionality of the dual construct was tested transiently in *N. benthamiana* using PDS as the silencing target, which resulted in the bleaching of the infiltrated area (see Supplementary Figure 7D).

An amiRNA silencing tool was also enabled with the creation of two special GBSperts, namely 5'FS and 3'FS, respectively. These GBSperts require noncanonical barcodes to allow the seamless assembly of 5'FS and 3'FS in the amiRNA precursor. The special categories are denoted as (12-13B) and (16B), respectively, where B indicates the four noncanonical flanking nucleotides (GTGA and TCTC, respectively). The standard (01-11) promoters without ATG and

the (17-21) terminators were used in the amiRNA design. The central region (14B-15B), containing a fragment of the gene target sequence, was constructed using gene-specific oligonucleotides, as described in Supplementary Figure 7B. In order to validate the proposed structure, *A. thaliana* PDS silencing was assayed using a gene target fragment which was formerly described by Yan et al. (2011) (Supplementary Figure 7E). The resulting amiRNA construct was transformed into *A.thaliana* yielding seedlings with the albino phenotype.

Finally in the hairpin RNA (hpRNA) structure, the regulatory regions lacking ATG are inserted as (01-11) parts. An intron from *S.lycopersicum* (SGN-U324070) was incorporated into the collection to serve as an (14-15) Intron GBpart. The inverted fragments of the target gene-of-interest can be cloned at positions (12-13) and at position (16).

#### *GW-GB adapter tool.*

A GW-GB adapter tool was incorporated into the GB2.0 collection in order to facilitate the transition between the Gateway (GW) and GB2.0 assembly methods (Supplementary Figure 8). GW-GB adapters are GBparts or GBSparts (e.g., a (12-16) GBSpart to adapt coding regions) made of a GW cassette flanked by attR1-attR2 sites and embedded inside the pUPD plasmid. As such, adapter vectors can be used directly as destination plasmids for GW entry clones flanked by attL1-attL2 sites. In this way, GW entry clones can be transferred individually or in bulk to the pUPD plasmid, and become ready-to-use GBSparts. Alternatively, the GW-GB adapter can be employed as an ordinary GBSpart to create a new multigene construct in a binary vector. Consequently, the resulting multigene construct becomes a GW destination vector containing an attR1-attR2 GW cassette, where GW entry clones can be inserted individually or in bulk. It should be noted that direct GW to GB2.0 adaptation does not remove internal enzyme target sites, therefore the efficiency of subsequent assembly reactions can lower.

### **GB collection and software tools**

When this manuscript was being written, our in-house GB collection contained more than 400 entries. As the collection grows, engineering is becoming increasingly easy and fast because, on occasion, the required GBparts, GBSparts and TUs are already domesticated and/or constructed. To efficiently handle this collection, we developed a web framework which hosts a GBdatabase and offers software tools to facilitate the assembly process. The GB2.0 website was implemented using Django, a Python web framework that supports rapid design and the development of web-based applications (Django, 2013). Object-relational database management system PostgreSQL was chosen to host our schema, which allowed the incorporation of the sequences of all the elements included in the collection. Additional relevant information on part identity, functionality and indexing is also provided.

Given the simplicity of the GB assembly rules, it was relatively straightforward to develop software tools that assist in GB2.0 assembly. We therefore developed a software package

comprising three programs, each program corresponding to one of the three basic processes in GB2.0 assembly. The first program, named GBDomesticator, assists the part adaptation process to the GB standard. It takes an input DNA sequence provided by the user, and it offers the best PCR strategy to remove internal enzyme target sites and to add flanking nucleotides to it according to the specified category. A second program, known as the TUassembler, takes GBparts and GBSparts from the database and simulates a multipartite assembly *in silico*. The TUassembler includes shortcuts to Frequently Used Structures assembly, as well as a free-hand option. Finally, a third program, namely BinaryAssembler, performs *in silico* binary assemblies between the composite parts stored in the GB database. BinaryAssembler offers the possibility of choosing the relative orientation of each member of the assembly. All three programs generate a detailed lab protocol to perform the domestication/assembly and to return a genbank formatted file containing the final domesticated/assembled sequence. The GB database and software tools are available at <http://www.gbcloning.org>

## DISCUSSION

The aim of this work is to provide a standard framework for DNA assembly in Plant Synthetic Biology. We, and others, realized that the modularity of the multipartite assembly based on type IIS enzymes offers a great opportunity for **standardization** by following a positional information scheme that resembles the grammar of a sentence in many natural languages. Indeed it is illustrative to conceive the transcriptional unit as a similar structure to a sentence, which is made up of hierarchically assembled elements like morphemes, words and phrases. It is also interesting to envision the whole engineering process as a way to imprint instructions using DNA strings. Therefore we, in concert with MoClo developers, propose a common grammar where the four nucleotide overhangs are pre-defined for each position within the transcriptional unit. Overhangs assignment is mainly arbitrary, but some decisions are made to make them scar-benign. For instance, the 12-13 boundary defining the beginning of CDS was designed to include the start codon, whereas the 13-14 boundary was made compatible with signal-peptide cleavage sites.

In our view, this new GB2.0 cloning scheme has a number of features which makes it a good candidate for a plant assembly standard. Many of those features are consubstantial to the Golden Gate system: very high efficiency, modularity and the ability to produce scar-benign assemblies. GB2.0 also incorporates the reusability and modularity of the GoldenBraid and MoClo systems and goes beyond them in that it provides a standardized framework, goes deep into the versatility and the minimalist design of the GoldenBraid loop, and incorporates new tools to assist cloning.

A major drawback of defining a standard is loss in **versatility** since no standard can cope with all custom design requirements. To deal with this problem, we incorporated an underlying non standard assembly level which makes full use of the newly designed pUPD vector. At this level, non standard GBpatches can be custom-designed for, e.g., scar-less assembly, by choosing the appropriate four nucleotide overhangs. GBpatches are assembled together into standard GBparts or GBSparts. We made full use of the GBpatch level for BiFC, amiRNA and antibody engineering. Other possible uses include promoter tinkering or non standard combinatorial assemblies within the CDS, as exemplified in the construction of customized TAL effectors (Weber et al., 2011; Li et al., 2012). Additionally, the GBpatch level is used for GBpart domestication; that is, for the removal of internal enzyme recognition sites. This feature is also enabled by the special design of the new entry vector pUPD, which introduces inversely oriented BsmBI sites into the GB cassette. This new design turns pUPD plasmid into a universal entry vector as the four nucleotides conferring part identity are not located in the entry vector as they are in previous designs (Weber et al., 2011). Instead in the present setup, the four-nucleotide “barcode” is incorporated into the primers used during initial part/patch isolation. As a toll, this strategy involves the requirement of longer PCR primers during initial part isolation.

This minor drawback is by far compensated by the simplicity introduced by the universal domesticator: in the absence of this solution, a minimum of eleven different entry vectors would be required to harbor the different categories in the GB grammar, along with an unaffordable amount of additional vectors to allow the formation of all the possible “phrasal” combinations.

The underlying GoldenBraid cloning pipeline has been substantially simplified in the GB2.0 version to reduce redundancy and to achieve a **minimalist design**. Figure 6 depicts the comparison of GB2.0 made with the previous GoldenBraid structure. Once again, most of the improvement achieved stems from the specific design of the new entry vector pUPD. First, the asymmetry of the cloning loop is corrected in GB2.0 with the introduction of a BtgZI site into the entry vector. BtgZI is a special enzyme that cuts 10 nucleotides away from its recognition site. This feature enables a dual release option for each GBpart: BsaI release allows cloning in  $\alpha$  destination vectors, whereas BtgZI release allows cloning in  $\Omega$  destination vectors. We noted that BtgZI/BsmBI assemblies are less efficient than BsaI ones. Despite this drawback, the ability to create new TUs in both destination vectors can save one cloning step, which therefore speeds up the construction of new multigene assemblies and opens up new possibilities for automation.

We also developed a number of tools to assist users in their engineering projects. First, we anticipated genetic designers’ needs by pre-arranging a number of FUS. Then, we populated our in-house collection with all the elements (GBparts, GBSparts and software tools) required to enable the Frequently Used Structures use. Finally, we assayed the functionality of newly developed elements using *in planta* assays. In certain cases, this implied an initial step toward part characterization. One of the hallmarks of Synthetic Biology is its ability to predict the behavior of a system based on the characteristics of its constitutive parts. We show herein that it is possible to infer the activity provided by a “promoter + terminator” pair from the activities that each individual element displays when separately assayed. The differences observed between the theoretical and experimental activity values fall within a narrow range which comes close to 0, with very few combinations showing deviations that are slightly above 2-fold ( $\pm$  0.3 in log values). This finding is important for engineering attempts which, as in complex metabolic engineering, require the combination of many different non coding parts to create large metabolic pathways, while avoiding the introduction of unstable repetitive regions into the genetic design. The promoter parts assayed herein reveal a wider range of activities than terminators. Nevertheless, we confirm that the use of strong terminators like TAtHSP18.2 can promote the promoter’s transcriptional activity, as previously described (Nagaya et al., 2009). It is interesting to note that most of the observed positive deviations result from the combinations involving CaMV35S-derived parts, suggesting a nonlinear behavior of the CaMV35S regulatory elements. We employed *N. benthamiana* transient expression and Luciferase/Renilla reporter system (Grentzmann et al., 1998) as a first step towards characterization of regulatory elements. This transient methodology is simple and accurate and therefore facilitates the

analysis. A more detailed characterization may need to include the developmental and tissue-specificity information obtained through stable plant transformation.

Both GB2.0 and the GB collections come into being with to an open-source vocation. We reinforced this point by developing a new set of GB-destination vectors based on open-source pCAMBIA binary vectors (Roberts et al., 1997; Chi-Ham et al., 2012). As we see it, the intellectual commons IP model is that which best suits the requirements for the free exchange of parts and modules in Plant Synthetic Biology (Oye and Wellhausen, 2010). Nevertheless, a number of issues, such as the IP of individual parts and the ability to freely distribute them, need to be addressed in a concerted manner. Undoubtedly, community effort made to create publicly available collections of synthetic parts will have an impact on the progress of this discipline.

Plant Synthetic Biology has the potential of bringing about a significant impact on crop production. Engineering enhanced abiotic stress tolerance for growth in marginal lands, turning C3 plants into C4 (Caemmerer et al., 2012), constructing whole-organism biosensors or sentinels (Antunes et al., 2011), engineering highly challenging metabolic routes (Farre et al., 2012), and combinations of these, are just some examples of high-impact goals with biotechnologists' reach. Also, it has not escaped our notice that the proposed grammar can be easily adopted by other non-plant systems as well. We believe that technologies like GB2.0, which enable the standardization and facilitate the characterization and exchange of genetic parts and modules, are important contributions for the achievement of the challenging biotechnology goals ahead.

## **MATERIAL AND METHODS**

### Strains and growth conditions

*Escherichia coli* DH5 $\alpha$  was used for cloning. *Agrobacterium tumefaciens* strain GV3101 was used for transient expression and transformation experiments. Both strains were grown in LB medium under agitation (200 rpm) at 37°C and 28°C, respectively. Ampicillin (50  $\mu\text{g ml}^{-1}$ ), kanamycin (50  $\mu\text{g ml}^{-1}$ ) and spectinomycin (100  $\mu\text{g ml}^{-1}$ ) were used for *E. coli* selection. Rifampicin, tetracycline and gentamicin were also used for *A. tumefaciens* selection at 50, 12.5 and 30  $\mu\text{g ml}^{-1}$ , respectively. XGal (0.5 mM) and IPTG (40  $\mu\text{g ml}^{-1}$ ) were used in LB agar plates for the white/blue selection of clones.

### Restriction-Ligation assembly reactions

Restriction-Ligation reactions were set up as described elsewhere (Sarrion-Perdigones et al., 2011) using BsaI, BsmBI, BtgZI or BbsI as restriction enzymes (New England Biolabs, Ipswich, MA, USA) and T4 Ligase (Promega, Madison, WI, USA). Reactions were set up in 25 or 50 cycles digestion/ligation reactions (2' at 37° C, 5' at 16° C), depending on assembly complexity. One  $\mu\text{l}$  of the reaction was transformed into *E. coli* DH5 $\alpha$  electrocompetent cells and positive clones were selected in solid media. Plasmid DNA was extracted using the E.Z.N.A. Plasmid Mini Kit I (Omega Bio-Tek, Norcross, GA, USA). Assemblies were confirmed by restriction analysis and sequencing.

### GBpart Domestication

GBparts and GBpatches were obtained by PCR amplification using suitable templates. The Phusion® High-Fidelity DNA Polymerase (ThermoScientific, Waltham, MA, USA) was used for amplification following the manufacturer's protocols. Primers smaller than 60 mers were purchased from Sigma-Aldrich (St. Louis, MO, USA). 60-mer or longer oligonucleotides were synthesized by IDTDNA (Coralville, IO, USA) by the Ultramer™ technology. Amplified bands were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and were quantified in a Nano Drop Spectrophotometer 2000. Then 40 ng of each amplicon and 75 ng of the domestication vector (pUPD) were mixed and incubated in a BsmBI restriction-ligation reaction. The pUPD sequence is deposited in the GBdatabase. Positive clones were selected in the ampicillin-, XGal- and IPTG -containing plates, and the correct assembly was confirmed by restriction analyses and sequencing. A description of the GBparts and GBSparts employed in this work is provided in Supplementary Table 1. The nucleotide sequence of all the GB parts in the collection is deposited in the GB database.

### pDGB Construction

Two pDGB series, pDGB1 and pDGB2, were constructed. pDGB1 is based on the pGreenII backbone (Hellens et al., 2000) and pDGB2 is based on pCAMBIA (Roberts et al., 1997). For pDGB construction, the backbone of each binary vector was divided into fragments (vector modules). The pDGB1 backbone comprised two fragments, whereas the pDGB2 backbone was divided into four modules given the presence of internal sites. To build vector modules, each fragment was amplified by PCR in a similar procedure to that described for GBparts and was cloned into a vector domestication plasmid (pVD) using a Bsal digestion-ligation reaction. The pVD vector was derived from pUPD; its sequence is deposited in GBdatabase. In addition to the backbone modules, a number of common modules were built: eight GB-cassettes ( $\alpha 1$ ,  $\alpha 1R$ ,  $\alpha 2$ ,  $\alpha 2R$ ,  $\Omega 1$ ,  $\Omega 1R$ ,  $\Omega 2$  and  $\Omega 2R$ ) and two fragments encoding spectinomycin and kanamycin resistance. To assemble each pDGB, a BbsI restriction-ligation reaction was performed by combining the modules of the vector backbone, the desired GB-cassette and appropriate antibiotic resistance.

#### *Nicotiana benthamiana* transient transformation

For the transient expression experiments, plasmids were transferred to *Agrobacterium tumefaciens* strain GV3101 by electroporation. Agroinfiltration was performed, as previously described (Wieland et al., 2006). Overnight-grown bacterial cultures were pelleted and resuspended in agroinfiltration medium (10 mM MES pH 5.6, 10 mM MgCl<sub>2</sub>, 200  $\mu$ M acetosyringone) to an optical density at 600 nm =0.5. Infiltrations were carried out using a needle-free syringe in leaves 2, 3 and 4 of 4–5 weeks old *Nicotiana benthamiana* plants (growing conditions: 24°C day/20°C night in a 16 h light/8 h dark cycle). Depending on the purpose of the experiments, leaves were harvested 3-5 days post-infiltration (d.p.i.) and examined for transgene expression.

#### *Arabidopsis thaliana* stable transformation.

*Arabidopsis thaliana* Col-0 accession plants were transformed by the floral-dip method (Clough and Bent, 1998). Seeds were sterilized-plated in plates of MS medium with 0.8% (w/v) agar and 1% (w/v) sucrose (growing conditions: 24°C day/20°C night in a 16h light/8h dark cycle). Transgenic lines were selected without antibiotic resistance as PDS silencing transformed lines showed the albino phenotype.

#### Renilla/Luciferase expression assays

In order to measure the activity of Renilla/Luciferase reporters (Grentzmann et al., 1998), 3 or 4 *Nicotiana benthamiana* leaves were agroinfiltrated following the above-described procedure. Leaves were harvested 3 d.p.i. *Firefly Luciferase* and *Renilla Luciferase* were assayed from 100-mg leaf extracts following the Dual-Glo Luciferase Assay System (Promega, Madison, WI, USA) standard protocol and were quantified with a GloMax 96 Microplate Luminometer (Promega, Madison, USA). The “ETA” of each region (ETA) was calculated in relation to the

Renilla/Luciferase values of a (01-12\_17-21) reference combination (PNos\_TNos), which was arbitrarily set as 1, according to the formulae:

$$ETA(01-12)_i = FLuc/RLuc [(01-12)_{i\_}(17-21)_{TNos}] / FLuc/RLuc [(01-12)_{PNos\_}(17-21)_{TNos}]$$

$$ETA(17-21)_j = FLuc/RLuc [(01-12)_{PNos\_}(17-21)_j] / FLuc/RLuc [(01-12)_{PNos\_}(17-21)_{TNos}],$$

where FLuc/RLuc [(01-12)<sub>i</sub>\_(17-21)<sub>j</sub>] refers to the ratio between the Firefly luciferase activity (FLuc) of a (01-12)<sub>i</sub>:Luciferase:(17-21)<sub>j</sub> construct and the Renilla luciferase activity (RLuc) of a 35S:Renilla:TNos internal standard construct. TTA was calculated for each cassette combination as the product of the individual ETA of the two regulatory regions, as follows:

$$TTA_{ij} = ETA(01-12)_i \times ETA(17-21)_j$$

Finally, the FLuc/RLuc of a number of cassette combinations was tested experimentally, and the ETA of each combination (ETA<sub>ij</sub>) was calculated with the formula:

$$ETA_{ij} = FLuc/RLuc [(01-12)_{i\_}(17-21)_j] / FLuc/RLuc [(01-12)_{PNos\_}(17-21)_{TNos}].$$

#### Glucocorticoid Receptor induction and Heat shock treatments

One-cm<sup>2</sup> disks from agroinfiltrated leaves were harvested at 3 d.p.i., placed in a 350 µl solution containing 5 to 20 µM Dexamethasone (Sigma Aldrich St. Louis, MO, USA) in 0.02% Tween-80 and incubated overnight in a growth chamber. *Firefly Luciferase* and *Renilla Luciferase* activities were measured after 24-hour treatment. For the heat shock treatments, 1 cm<sup>2</sup> of 3 d.p.i. leaves were placed in 350 µl water at 37°C for 2h. Samples were collected at 3 h and 14 h after treatment.

## **ACKNOWLEDGMENTS**

We wish to particularly thank Dr. Marillonnet for his efforts in agreeing on a common standard for MoClo and GB2.0. Many people provided us with material and knowledge, which helped us to build our GBCollection. From our Research Centre (IBMCP, Spain), Dr. Blázquez and Dr. Alabadí shared their knowledge on *A. thaliana* promoters, Dr. Ferrando shared the BiFC vectors, Dr. Ferrandiz helped with the protein-protein interaction examples, Dr. Carbonell and Dr. Urbez helped with the amiRNA work, while Dr. Darós provided assistance with the fluorescent proteins and silencing suppressors. HSP70B was provided by Prof. Maule (John Innes Center, UK). Dr. Felippes and Dr. Wiegel shared detailed information on the MIGS design. This work has been supported by the Spanish Ministry of Economy and Competitiveness: Grant BIO2010-15384. A. Sarrion-Perdigones is the recipient of a Research Personnel in Training (FPI) fellowship and M. Vázquez-Vilar is the recipient of a Junta de Ampliación de Estudios (JAE) fellowship.

## FIGURE LEGENDS

**Figure 1. Analogies between GB2.0 and English grammar** (A) GB2.0 elements can be compared with those of a natural language. In English grammar (left), morphemes are joined together to make words; words are combined together to make phrases and sentences, which are further joined to make a composition. In GB2.0 (right), the simplest units are GBpatches, used to build any of the 11 standard GBparts. GBpatches can be also combined in GBSparts to facilitate cloning (e.g., a whole promoter). GBparts and GBSparts are combined in a multipartite reaction to build TUs, which can be used for plant transformation, or can be reused and combined with other TUs to build multigene modules. (B) The flow chart of the GB assembly steps. It starts with the GB domestication of GBpatches into GBparts or GBSparts; GBparts are multipartitely combined to build up TUs; finally, TUs are binarily assembled to build modules and multigene constructs.

**Figure 2. The complete GB2.0 grammar and its most Frequently Used Structures.** (A) A schematic overview of a TU structure where the 11 standard GB classes are depicted: 01, 02 and 03 GBparts form the 5' non transcribed region (5'NT); position 11 is the 5' untranslated region of mRNA (5' UTR); 12 is a linker region; 13, 14, 15 and 16 (TL1 to TL4) are four divisions of the translated region; 17 is the 3' untranslated region of mRNA (3' UTR) and 21 is the 3' non transcribed region of the TU (3'NT). (B) Frequently-used structures (FUS) for the protein-coding TUs. The elements forming each Frequently Used Structure and the class that they belong to are depicted. (C) Frequently Used Structure for RNA silencing, including artificial micro RNA (amiRNA), hairpin RNA (hpRNA) and trans-acting small interfering RNAs (tasiRNA). 5'NT and 3'NT indicate the 5' and 3' non transcribed regions of the TU; 5'UTR and 3'UTR are the 5' and 3' untranslated regions of mRNA; LINK represents a region between the 5'UTR and the coding sequence where tags or fused proteins can be placed; PROM is a promoter; CDS is the coding DNA sequence; TER represents the terminator; SP is signal peptide; NT and CT are N- or C-Terminal tags or fusion proteins; OP is a promoter operator; minPROM is a minimal promoter. 5'FS and 3'FS indicate the flanking sequences of the amiRNA precursor sequences; TARGET represents the region of the amiRNA structure comprising the loop and the complementary target sequences; GOI and IOG are the fragments of the gene of interest in an inverted orientation; INT is the intron for hpRNA processing; mir173 represents the mir173 target site for tasiRNA processing; fGOI indicates the fragment gene of interest to be silenced.

**Figure 3. Standardized domestication of GBparts.** GBparts are domesticated by amplifying the desired sequence with standard GBprimers (GB.F and GB.R). GBprimers include approximately 20 nucleotides of the GSP (Gene-Specific Primer) and a tail region that includes a BsmBI recognition site, the cleavage site for cloning into pUPD and the four nucleotide barcode (1234 and 5678 in the figure). The amplified DNA part is cloned into pUPD in a restriction-ligation reaction, with BsmBI as the restriction enzyme. The resulting GBpart is

cleavable by Bsal and BtgZI to produce 1234 and 5678 flanking overhangs. BsmBI recognition sequences are depicted in orange in the DNA sequence; Bsal and BtgZI are labeled in red and blue, respectively. Enzyme cleavage sites are boxed.

**Figure 4. GB2.0 cassettes and assembly rules. (A)** GB2.0 cassettes and their comparison with previous GoldenBraid version. GBcassettes comprise a LacZ selection cassette flanked by four Type IIs restriction sites (Bsal, BsmBI) positioned in inversed orientations. Previous GoldenBraid plasmid kit comprised four destination plasmids, two in each assembly level. GB2.0 incorporates four additional plasmids that permit the assembly of transcriptional units in reverse orientation using the same GBparts. Additionally, the six 4 nt barcodes of GoldenBraid (A, B, C, 1, 2 and 3) collapsed in only three GB2.0 barcodes, where  $A \equiv 1$ ,  $B \equiv 2$ ,  $C \equiv 3$ . This special design feature permits GBparts to be directly assembled in both level plasmids. Finally, GB2.0 plasmids incorporate distinctive restriction sites flanking the GBcassette as watermarks for plasmid identification. Bsal cleavage sequences are boxed in red, BsmBI cleavage sequences are boxed in orange and sites where both enzymes can digest are boxed in green. The watermark restriction sites are underlined. (B) Rules for Multipartite assemblies. The pUPD elements represent each GBparts and GBSparts that conforms a grammatically correct TU, pDGB $\Omega$ <sub>i</sub> is any level  $\Omega$  destination vector, pDGB $\alpha$ <sub>i</sub> is any level  $\alpha$  destination vector, and pEGB $\Omega$ <sub>i</sub>(X) and pEGB $\alpha$ <sub>i</sub>(X) are the resulting expression plasmids harboring a well-constructed transcriptional unit X. (C) Rules for binary assemblies. (X<sub>i</sub>) and (X<sub>j</sub>) are composite parts assembled using the multipartite assembly option; (X<sub>i</sub>+X<sub>j</sub>) is a composite part of (X<sub>i</sub>) and (X<sub>j</sub>) that follows the same assembly rules than (X<sub>i</sub>) and (X<sub>j</sub>); pEGB $\alpha$ <sub>1</sub>(X), pEGB $\alpha$ <sub>2</sub>(X), pEGB $\Omega$ <sub>1</sub>(X) and pEGB $\Omega$ <sub>2</sub>(X) are expression plasmids hosting a composite part X; and pDGB $\Omega$ <sub>1</sub>, pDGB $\Omega$ <sub>2</sub>, pDGB $\alpha$ <sub>1</sub> and pDGB $\alpha$ <sub>2</sub> are destination plasmids hosting a LacZ cassette.

**Figure 5. Characterization of regulatory regions for basic expression cassettes. (A)** Constructs for ETA quantification. The promoter (01-12)<sub>i</sub>-(17-21)<sub>TN $\alpha$ s</sub> constructs comprise a first TU with the (01-12) promoter of interest, the firefly luciferase and the Nopaline syntase terminator, followed by the Renilla reference module (upper row). For the (01-12)<sub>TN $\alpha$ s</sub>-(17-21)<sub>j</sub> terminator constructs, the first TU comprises the (17-21) terminator of interest, the firefly luciferase and the Nopaline syntase promoter (central row). For activity normalization, the PNos:Luciferase:TNos construct combined with the Renilla reference module was used (lower row). (B) The ETA of the promoter regions in (01-12)<sub>i</sub>-(17-21)<sub>TN $\alpha$ s</sub> constructs was determined as the Firefly (FLuc)/ Renilla (RLuc) Luciferase activity ratios of each construct normalized with the equivalent ratio of the PNos:Luciferase:TNos construct. Error bars represent the SD of at least three replicates. (C) The ETA of terminator regions in the (01-12)<sub>PNos</sub>-(17-21)<sub>j</sub> constructs was estimated as described in B. (D) Scheme of the combinatorial promoter\_terminator constructs comprising a first TU with a (01-12) promoter, the firefly luciferase and a (17-21) terminator, and combined with the Renilla reference module. (E) Correspondence between the experimental (ETA) and theoretical (TTA) activity data in the combinatorial constructs. The logarithm of the ratios between the ETA and TTA values for 34 experimental promoter/terminator combinations

is plotted. Bars represent average values  $\pm$  SD. PNos, nopaline syntase promoter; TNos, nopaline syntase terminator; P35s, CaMV35s promoter; p19, TBSV silencing suppressor.

**Figure 6. Goldenbraid versus GB2.0.** The previous GoldenBraid version had an asymmetric assembly flow (left). GBparts incorporated either the Bsal or the BsmBI releasable overhangs. Bsal-released GBparts were incorporated into the GoldenBraid cloning loop through level- $\alpha$  vectors, whereas the BsmBI GBparts were used to build composite parts through the level- $\Omega$  entry point. In the new GB2.0 symmetric design, the same GBparts can be incorporated into level- $\alpha$  plasmids by a Bsal restriction/ligation reaction or into level- $\Omega$  vectors by a mixed BsmBI/BtgZI reaction. Other differences between the previous GoldenBraid version and GB2.0 are also listed in the figure.

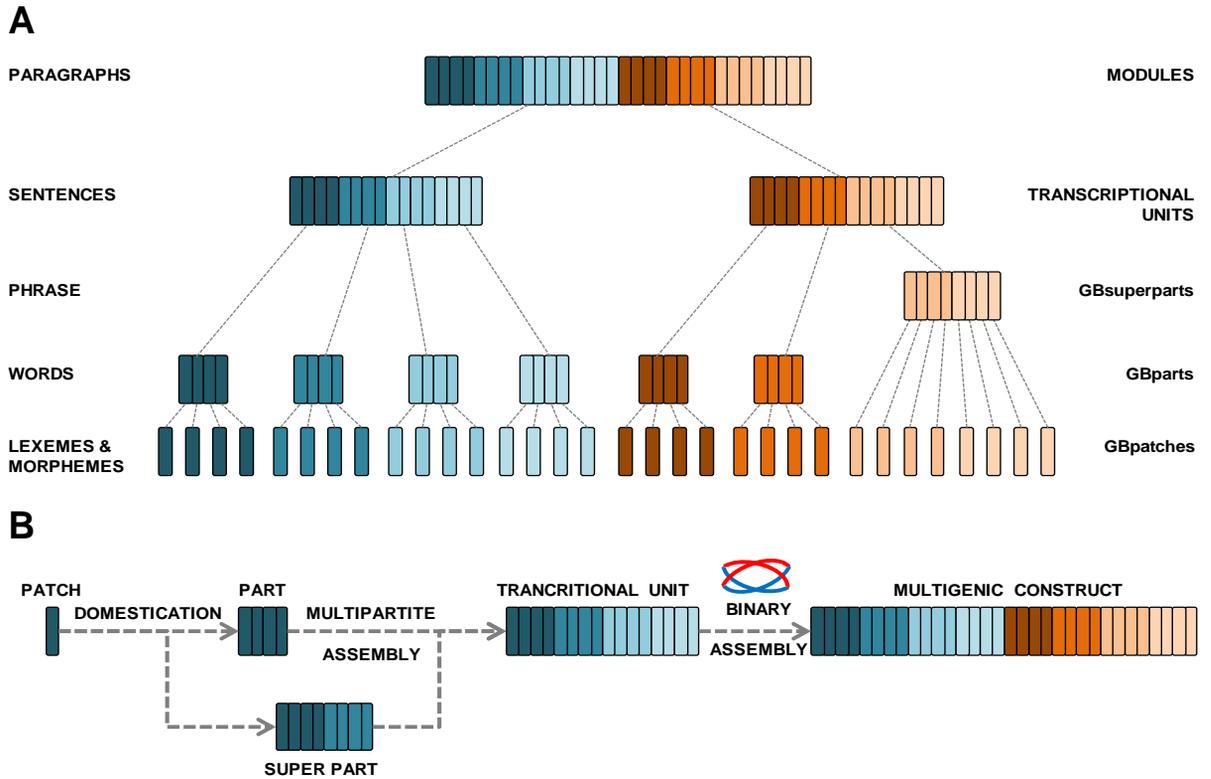
## REFERENCES

- Antunes MS, Morey KJ, Smith JJ, Albrecht KD, Bowen TA, Zdunek JK, Troupe JF, Cuneo MJ, Webb CT, Hellinga HW, Medford JI** (2011) Programmable ligand detection system in plants through a synthetic signal transduction pathway. *PLoS One* **6**: e16292
- Aslanidis C, de Jong PJ** (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**: 6069-6074
- Belda-Palazon B, Ruiz L, Marti E, Tarraga S, Tiburcio AF, Culianez F, Farras R, Carrasco P, Ferrando A** (2012) Aminopropyltransferases involved in polyamine biosynthesis localize preferentially in the nucleus of plant cells. *PLoS One* **7**: e46907
- Caemmerer Sv, Quick W, Furbank R** (2012) The Development of C4 Rice: Current Progress and Future Challenges. *Science* **336**: 2
- Clough SJ, Bent AF** (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743
- Check E** (2005) Synthetic biology: Designs on life. *Nature* **438**: 417 - 418
- Chen QJ, Xie M, Ma XX, Dong L, Chen J, Wang XC** (2010) MISSA is a highly efficient in vivo DNA assembly method for plant multiple-gene transformation. *Plant Physiol* **153**: 41-51
- Chen QJ, Zhou HM, Chen J, Wang XC** (2006) A Gateway-based platform for multigene plant transformation. *Plant Mol Biol* **62**: 927 - 936
- Chi-Ham CL, Boettiger S, Figueroa-Balderas R, Bird S, Geoola JN, Zamora P, Alandete-Saez M, Bennett AB** (2012) An intellectual property sharing initiative in agricultural biotechnology: development of broadly accessible technologies for plant transformation. *Plant Biotechnol J* **10**: 501-510
- Dafny-Yelin M, Tzfira T** (2007) Delivery of multiple transgenes to plant cells. *Plant Physiol* **145**: 1118-1128
- Django** (2013) Django (Version 1.5) Retrieved from <http://djangoproject.com> *In*. Django Software Foundation
- Ellis T, Adie T, Baldwin GS** (2011) DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol* **3**: 109-118
- Engler C, Gruetzner R, Kandzia R, Marillonnet S** (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS One* **4**: e5553
- Engler C, Kandzia R, Marillonnet S** (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One* **3**: e3647
- Estornell LH, Orzaez D, Lopez-Pena L, Pineda B, Anton MT, Moreno V, Granell A** (2009) A multisite gateway-based toolkit for targeted gene expression and hairpin RNA silencing in tomato fruits. *Plant Biotechnol J* **7**: 298-309
- Farre G, Naqvi S, Sanahuja G, Bai C, Zorrilla-Lopez U, Rivera SM, Canela R, Sandman G, Twyman RM, Capell T, Zhu C, Christou P** (2012) Combinatorial genetic transformation of cereals and the creation of metabolic libraries for the carotenoid pathway. *Methods Mol Biol* **847**: 419-435

- Felippes FF, Wang JW, Weigel D** (2012) MIGS: miRNA-induced gene silencing. *Plant J* **70**: 541-547
- Fujisawa M, Takita E, Harada H, Sakurai N, Suzuki H, Ohyama K, Shibata D, Misawa N** (2009) Pathway engineering of *Brassica napus* seeds using multiple key enzyme genes involved in ketocarotenoid formation. *J Exp Bot* **60**: 1319-1332
- Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA** (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res* **35**: e55
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO** (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Meth* **6**: 343-345
- Greutzmann G, Ingram JA, Kelly PJ, Gesteland RF, Atkins JF** (1998) A dual-luciferase reporter system for studying recoding signals. *RNA* **4**: 479-486
- Haseloff J, Ajioka J** (2009) Synthetic biology: history, challenges and prospects. *J R Soc Interface* **6 Suppl 4**: S389-391
- Hellens RP, Edwards EA, Leyland NR, Bean S, Mullineaux PM** (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. *Plant Mol Biol* **42**: 819-832
- Karimi M, Bleys A, Vanderhaeghen R, Hilson P** (2007) Building blocks for plant gene assembly. *Plant Physiology* **145**: 1183-1191
- Karimi M, Inze D, Depicker A** (2002) GATEWAY vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci* **7**: 193-195.
- Knight TF** (2003) Idempotent Vector Design for Standard Assembly of BioBricks. Tech. rep., MIT Synthetic Biology Working Group Technical Reports
- Li L, Piatek MJ, Atef A, Piatek A, Wibowo A, Fang X, Sabir JS, Zhu JK, Mahfouz MM** (2012) Rapid and highly efficient construction of TALE-based transcriptional regulators and nucleases for genome modification. *Plant Mol Biol* **78**: 407-416
- Lin L, Liu Y-G, Xu X, Li B** (2003) Efficient linking and transfer of multiple genes by a multigene assembly and transformation vector system. *Proc Natl Acad Sci* **100**: 5962-5967
- Nagaya S, Kawamura K, Shinmyo A, Kato K** (2009) The HSP terminator of *Arabidopsis thaliana* increases gene expression in plant cells. *Plant Cell Physiol* **51**: 328-332
- Oye KA, Wellhausen R** (2010) The Intellectual Commons and Property in Synthetic Biology. *In* M Schmidt, ed, *Synthetic Biology: The technoscience and its societal consequences*. Springer Netherlands, pp 121-140
- Roberts C, Rajagopal,S., Smith,L.M., Nguyen,T.A., Yang,W.,, Nugrohu S, Ravi,K.S., Vijayachandra,K., Harcourt,R.L.,, Dransfield L, Desamero,N., Slamet,I., Hadjukiewicz,P., Svab,Z.,, Maliga P, Mayer,J.E., Keese,P.K., Kilian,A. and Jefferson,R.A.** (1997) A comprehensive set of modular vectors for advanced manipulations and efficient transformation of plants. pCAMBIA Vector Release Manual

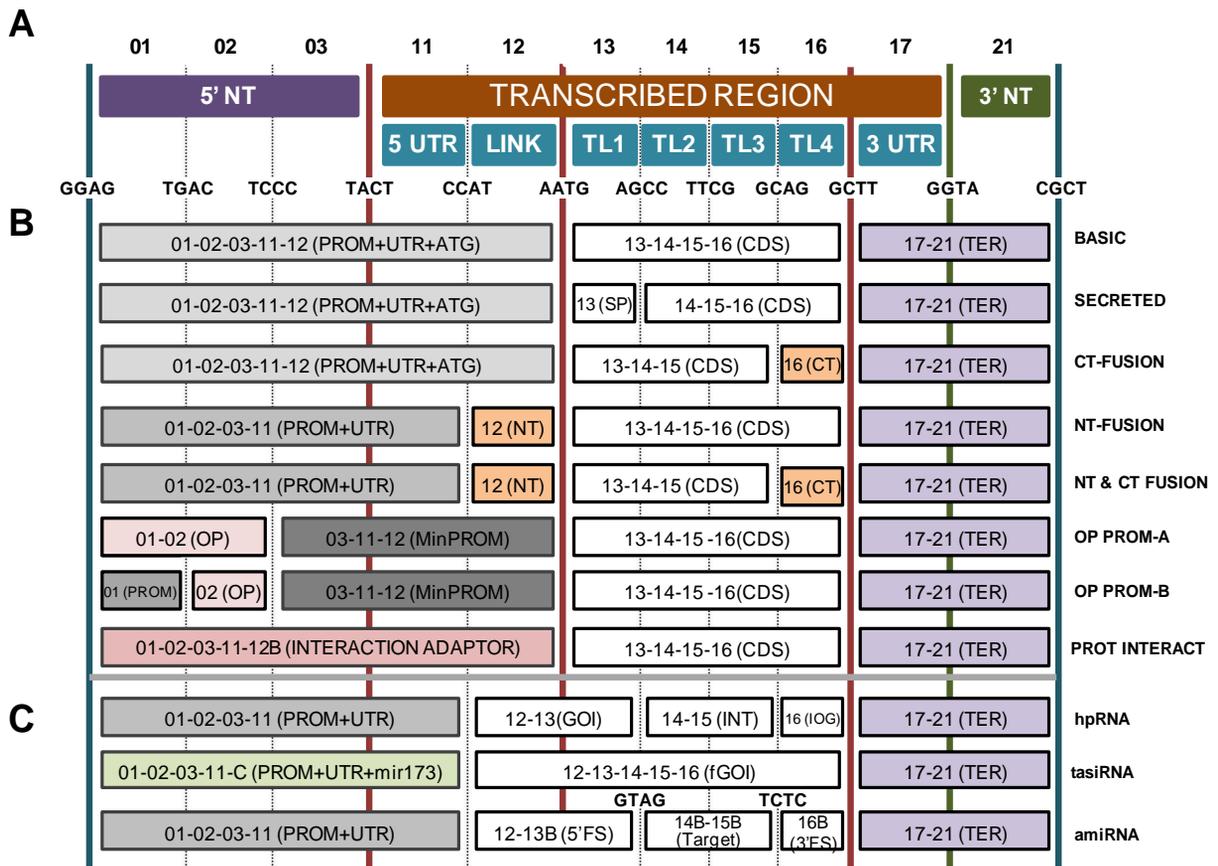
- Sarrion-Perdigones A, Falconi EE, Zandalinas SI, Juarez P, Fernandez-del-Carmen A, Granell A, Orzaez D** (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One* **6**: e21622
- Vemanna RS, Chandrashekar BK, Hanumantha Rao HM, Sathyanarayanagupta SK, Sarangi KS, Nataraja KN, Udayakumar M** (2013) A modified MultiSite gateway cloning strategy for consolidation of genes in plants. *Mol Biotechnol* **53**: 129-138
- Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S** (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**: e16765
- Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S** (2011) Assembly of designer TAL effectors by Golden Gate cloning. *PLoS One* **6**: e19722
- Wieland WH, Lammers A, Schots A, Orzaez DV** (2006) Plant expression of chicken secretory antibodies derived from combinatorial libraries. *J Biotechnol* **122**: 382-391
- Yan H, Deng X, Cao Y, Huang J, Ma L, Zhao B** (2011) A novel approach for the construction of plant amiRNA expression vectors. *J Biotechnol* **151**: 9-14
- Zeevi V, Liang Z, Arieli U, Tzfira T** (2012) Zinc finger nuclease and homing endonuclease-mediated assembly of multigene plant transformation vectors. *Plant Physiol* **158**: 132-144

# Figure 1



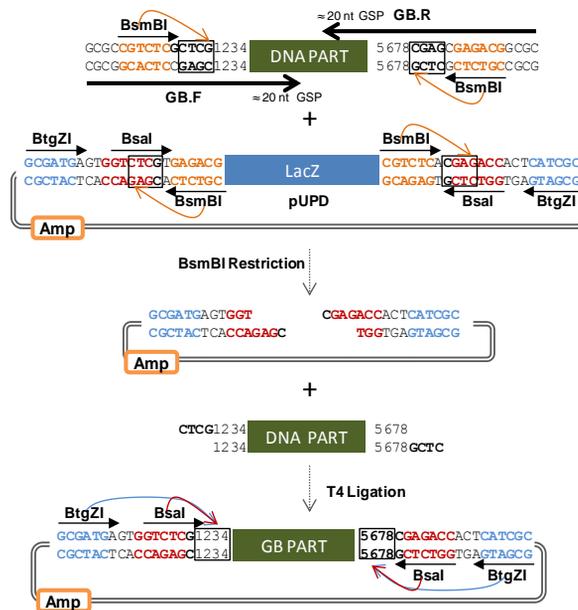
**Figure 1. Analogies between GB2.0 and English grammar** (A) GB2.0 elements can be compared with those of a natural language. In English grammar (left), morphemes are joined together to make words; words are combined together to make phrases and sentences, which are further joined to make a composition. In GB2.0 (right), the simplest units are GBpatches, used to build any of the 11 standard GBparts. GBpatches can be also combined in GBSparts to facilitate cloning (e.g., a whole promoter). GBparts and GBSparts are combined in a multipartite reaction to build TUs, which can be used for plant transformation, or can be reused and combined with other TUs to build multigene modules. (B) The flow chart of the GB assembly steps. It starts with the GB domestication of GBpatches into GBparts or GBSparts; GBparts are multipartitely combined to build up TUs; finally, TUs are binarily assembled to build modules and multigene constructs.

# Figure 2



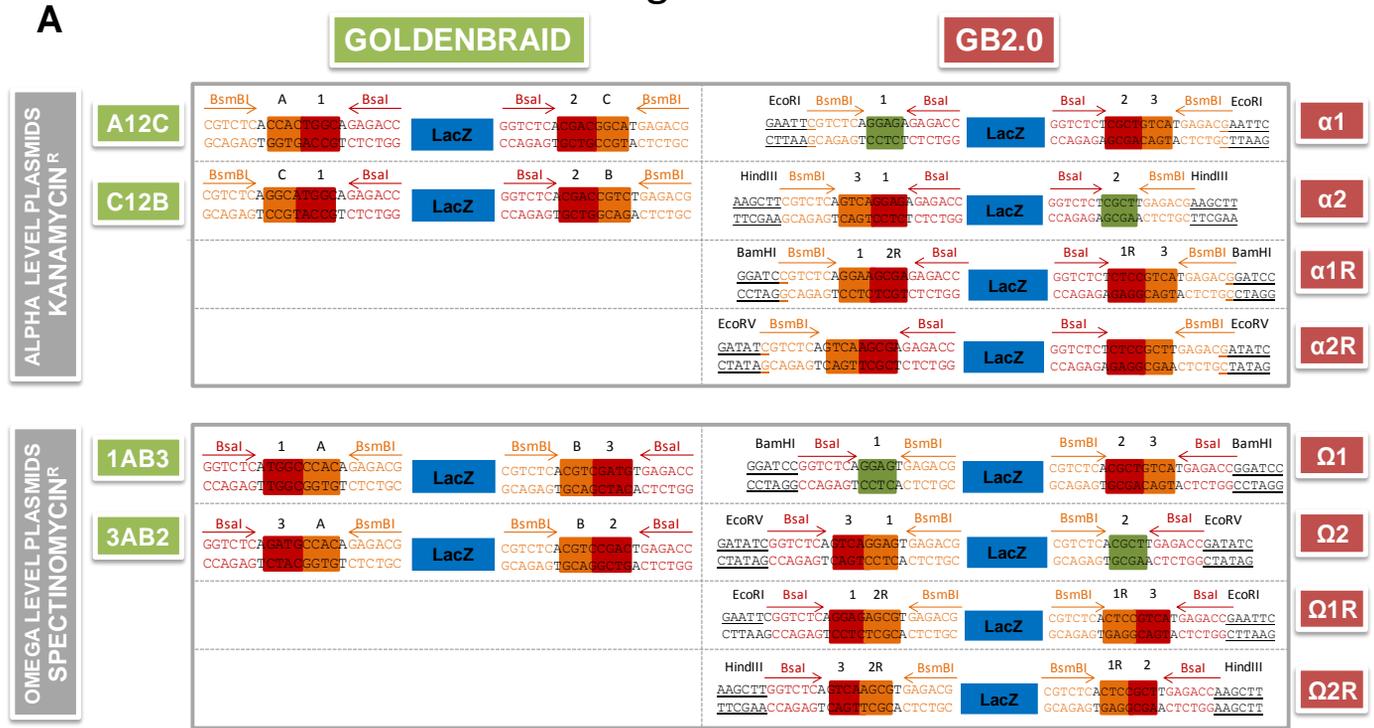
**Figure 2. The complete GB2.0 grammar and its most Frequently Used Structures.** (A) A schematic overview of a TU structure where the 11 standard GB classes are depicted: 01, 02 and 03 GBparts form the 5' non transcribed region (5'NT); position 11 is the 5' untranslated region of mRNA (5' UTR); 12 is a linker region; 13, 14, 15 and 16 (TL1 to TL4) are four divisions of the translated region; 17 is the 3' untranslated region of mRNA (3' UTR) and 21 is the 3' non transcribed region of the TU (3'NT). (B) Frequently-used structures (FUS) for the protein-coding TUs. The elements forming each FUS and the class that they belong to are depicted. (C) FUS for RNA silencing, including artificial micro RNA (amiRNA), hairpin RNA (hpRNA) and trans-acting small interfering RNAs (tasiRNA). 5'NT and 3'NT indicate the 5' and 3' non transcribed regions of the TU; 5'UTR and 3'UTR are the 5' and 3' untranslated regions of mRNA; LINK represents a region between the 5'UTR and the coding sequence where tags or fused proteins can be placed; PROM is a promoter; CDS is the coding DNA sequence; TER represents the terminator; SP is signal peptide; NT and CT are N- or C-Terminal tags or fusion proteins; OP is a promoter operator; minPROM is a minimal promoter. 5'FS and 3'FS indicate the flanking sequences of the amiRNA precursor sequences; TARGET represents the region of the amiRNA structure comprising the loop and the complementary target sequences; GOI and IOG are the fragments of the gene of interest in an inverted orientation; INT is the intron for hpRNA processing; mir173 represents the mir173 target site for tasiRNA processing; fGOI indicates the fragment gene of interest to be silenced.

### Figure 3



**Figure 3. Standardized domestication of GBparts.** GBparts are domesticated by amplifying the desired sequence with standard GBprimers (GB.F and GB.R). GBprimers include approximately 20 nucleotides of the GSP (Gene-Specific Primer) and a tail region that includes a BsmBI recognition site, the cleavage site for cloning into pUPD and the four nucleotide barcode (1234 and 5678 in the figure). The amplified DNA part is cloned into pUPD in a restriction-ligation reaction, with BsmBI as the restriction enzyme. The resulting GBpart is cleavable by Bsal and BtgZI to produce 1234 and 5678 flanking overhangs. BsmBI recognition sequences are depicted in orange in the DNA sequence; Bsal and BtgZI are labeled in red and blue, respectively. Enzyme cleavage sites are boxed.

Figure 4



**B RULES FOR MULTIPARTITE ASSEMBLY**

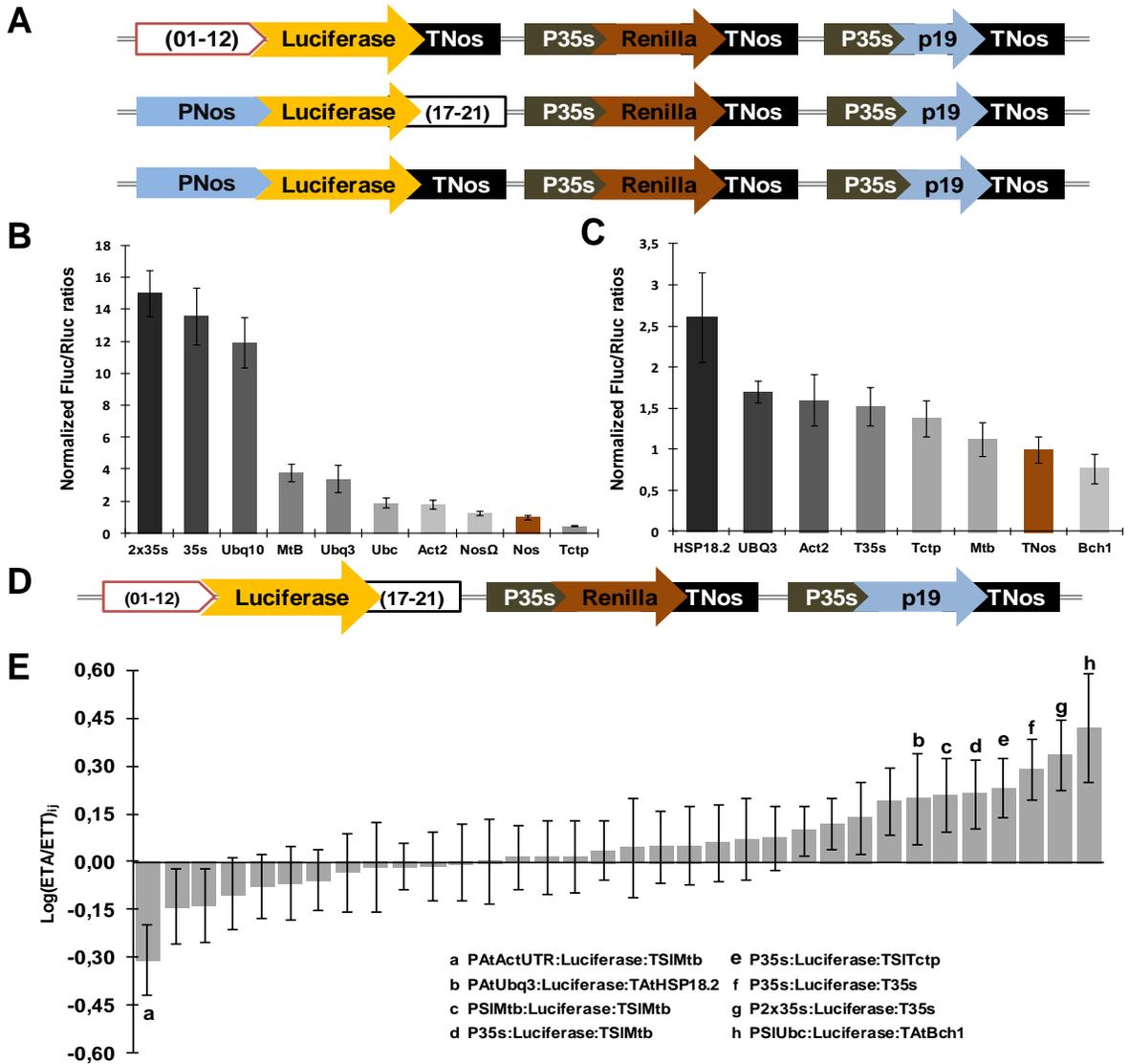
$pUPD(01-a)... + pUPD(c-d)... + pUPD(f-21) + pDGB\alpha i + Bsal + ligase = pEGB\alpha i (X)$   
 $pUPD (01-a)... + pUPD (c-d)... + pUPD (f-21) + pDGB\Omega i + BsmBI + BtgZI + ligase = pEGB\Omega i (X)$

**C RULES FOR BINARY ASSEMBLY**

$pE\alpha 1(X_i) + pE\alpha 2(X_j) + pDGB\Omega 1 + BsmBI + ligase = pEGB\Omega 1(X_i+X_j)$   
 $pE\alpha 1(X_i) + pE\alpha 2(X_j) + pDGB\Omega 2 + BsmBI + ligase = pEGB\Omega 2(X_i+X_j)$   
 $pE\Omega 1 (X_i) + pE\Omega 2(X_j) + pDGB\alpha 1 + Bsal + ligase = pEGB\alpha 1(X_i+X_j)$   
 $pE\Omega 1 (X_i) + pE\Omega 2(X_j) + pDGB\alpha 2 + Bsal + ligase = pEGB\alpha 2(X_i+X_j)$

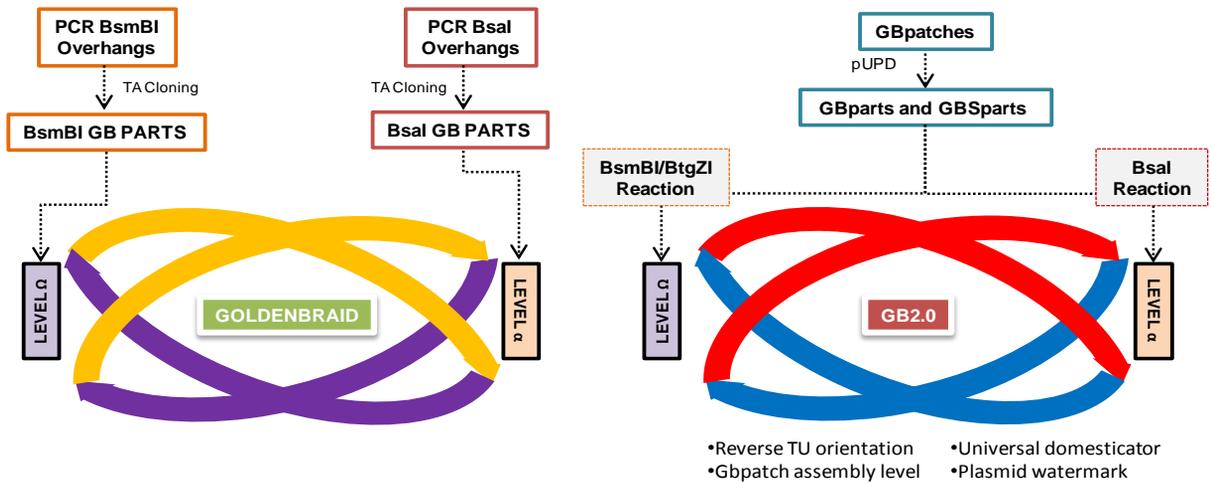
**Figure 4. GB2.0 cassettes and assembly rules. (A)** GB2.0 cassettes and their comparison with previous GoldenBraid version. GBcassettes comprise a LacZ selection cassette flanked by four Type IIIs restriction sites (Bsal, BsmBI) positioned in inversed orientations. Previous GoldenBraid plasmid kit comprised four destination plasmids, two in each assembly level. GB2.0 incorporates four additional plasmids that permit the assembly of transcriptional units in reverse orientation using the same GBparts. Additionally, the six 4 nt barcodes of GoldenBraid (A, B, C, 1, 2 and 3) collapsed in only three GB2.0 barcodes, where  $A \equiv 1$ ,  $B \equiv 2$ ,  $C \equiv 3$ . This special design feature permits GBparts to be directly assembled in both level plasmids. Finally, GB2.0 plasmids incorporate distinctive restriction sites flanking the GBcassete as watermarks for plasmid identification. Bsal cleavage sequences are boxed in red, BsmBI cleavage sequences are boxed in orange and sites where both enzymes can digest are boxed in green. The watermark restriction sites are underlined. (B) Rules for Multipartite assemblies. The pUPD elements represent each GBparts and GBSparts that conforms a grammatically correct TU, pDGB $\Omega$ i is any level  $\Omega$  destination vector, pDGB $\alpha$ i is any level  $\alpha$  destination vector, and pEGB $\Omega$ i (X) and pEGB $\alpha$ i (X) are the resulting expression plasmids harboring a well-constructed transcriptional unit X. (C) Rules for binary assemblies. (X<sub>i</sub>) and (X<sub>j</sub>) are composite parts assembled using the multipartite assembly option; (X<sub>i</sub>+X<sub>j</sub>) is a composite part of (X<sub>i</sub>) and (X<sub>j</sub>) that follows the same assembly rules than (X<sub>i</sub>) and (X<sub>j</sub>); pEGB $\alpha$ 1(X), pEGB $\alpha$ 2(X), pEGB $\Omega$ 1(X) and pEGB $\Omega$ 2(X) are expression plasmids hosting a composite part X; and pDGB $\Omega$ 1, pDGB $\Omega$ 2, pDGB $\alpha$ 1 and pDGB $\alpha$ 2 are destination plasmids hosting a LacZ cassette.

Figure 5



**Figure 5. Characterization of regulatory regions for basic expression cassettes.** (A) Constructs for ETA quantification. The promoter  $(01-12)_{i-}(17-21)_{TNos}$  constructs comprise a first TU with the  $(01-12)$  promoter of interest, the firefly luciferase and the Nopaline syntase terminator, followed by the Renilla reference module (upper row). For the  $(01-12)_{TNos-}(17-21)_j$  terminator constructs, the first TU comprises the  $(17-21)$  terminator of interest, the firefly luciferase and the Nopaline syntase promoter (central row). For activity normalization, the PNos:Luciferase:TNos construct combined with the Renilla reference module was used (lower row). (B) The ETA of the promoter regions in  $(01-12)_{i-}(17-21)_{TNos}$  constructs was determined as the Firefly (FLuc)/ Renilla (RLuc) Luciferase activity ratios of each construct normalized with the equivalent ratio of the PNos:Luciferase:TNos construct. Error bars represent the SD of at least three replicates. (C) The ETA of terminator regions in the  $(01-12)_{PNos-}(17-21)_j$  constructs was estimated as described in B. (D) Scheme of the combinatorial promoter\_terminator constructs comprising a first TU with a  $(01-12)$  promoter, the firefly luciferase and a  $(17-21)$  terminator, and combined with the Renilla reference module. (E) Correspondence between the experimental (ETA) and theoretical (TTA) activity data in the combinatorial constructs. The logarithm of the ratios between the ETA and TTA values for 34 experimental promoter/terminator combinations is plotted. Bars represent average values  $\pm$  SD. PNos, nopaline syntase promoter; TNos, nopaline syntase terminator; P35s, CaMV35s promoter; p19, TBSV silencing suppressor.

Figure 6



**Figure 6. Goldenbraid versus GB2.0.** The previous GoldenBraid version had an asymmetric assembly flow (left). GBparts incorporated either the Bsal or the BsmBI releasable overhangs. Bsal-released GBparts were incorporated into the GoldenBraid cloning loop through level- $\alpha$  vectors, whereas the BsmBI GBparts were used to build composite parts through the level- $\Omega$  entry point. In the new GB2.0 symmetric design, the same GBparts can be incorporated into level- $\alpha$  plasmids by a Bsal restriction/ligation reaction or into level- $\Omega$  vectors by a mixed BsmBI/BtgZI reaction. Other differences between the previous GoldenBraid version and GB2.0 are also listed in the figure.