

# In vivo evolution of metabolic pathways

## Assembling old parts to build novel and functional structures

Alejandro Luque\*, Sarra C Sebai, Vincent Sauveplane, Odile Ramaen, and Rudy Pandjaitan  
EVIAGENICS; Villejuif, France

**Keywords:** OXA genes, phenylpropanoids, flavonoids, homeologous/homologous recombination, DNA repair deficiency, in vivo evolution, mosaic genes and pathways, metabolic engineering, yeast

© Alejandro Luque, Sarra C Sebai, Vincent Sauveplane, Odile Ramaen, and Rudy Pandjaitan

\*Correspondence to: Alejandro Luque; Email: luque@eviagenics.com

Submitted: 07/28/2014

Revised: 08/03/2014

Accepted: 08/05/2014

<http://dx.doi.org/10.4161/bioe.34347>

Commentary to: Luque, AE, Sebai, S, Santiago-Schübel, B, Le Coz, Y, Jenot, D, Ramaen, O, Sauveplane, V and Pandjaitan, R (2014). In vivo evolution of metabolic pathways by homeologous recombination in mitotic cells. *Metab. Eng.* 23:123-135. <http://dx.doi.org/10.1016/j.jymben.2014.02.010>

This is an Open Access article distributed under the terms of the Creative Commons Attribution-Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

**I**n our recent article “In vivo evolution of metabolic pathways by homeologous recombination in mitotic cells” we proposed a useful alternative to directed evolution methods that permits the generation of yeast cell libraries containing recombinant metabolic pathways from counterpart genes. The methodology was applied to generate single mosaic genes and intragenic mosaic pathways. We used flavonoid metabolism genes as a working model to assemble and express evolved pathways in DNA repair deficient cells. The present commentary revises the principles of gene and pathway mosaicism and explores the scope and perspectives of our results as an additional tool for synthetic biology.

### Introduction

Nowadays the replacement of petrol-based expensive industrial processes by natural proceedings less offensive to the environment has become an indisputable aim for applied life sciences. Last decades can be considered in that sense as an intricate circuit on which both, academia and industry have competed in re-designing and optimizing all kind of biological systems and bioprocesses for useful purposes. The race is far from being over if one considers what the goal implies. In fact, technological trends and the increased knowledge in bioengineering condition new questions and challenges at the same time that emerging alternatives define new branching points on the road. From nutraceutical to biofuels and therapeutic drugs, synthetic biology—defined as the field of bioengineering by which cells are re-designed and re-programmed with new

genetic information to acquire new behaviors—specially has played a pivotal pilot role in the development of this “green optimization” race during past 20 y. Recent reviews highlight trends on the topic, timeline, and perspectives.<sup>1-3</sup>

More than single enzymatic steps, the synthesis of biomolecules of interest often demands a cascade of orchestrated reactions, such as metabolic pathways or networks occurring in nature, which need to be provided and adapted to the host cell. In that way bioengineers assemble genetic parts, the BioBricks, obtained from some organisms to build up new cell producers in the best metabolic conditions.<sup>4,5</sup> However, the assembly of existing parts just by combining them not always results in an optimal functionality or appropriated synthesis. For example, the target cell must be adapted as a metabolic chassis able to sustain the genetic program it receives. Cell stability, toxicity, and availability of synthesized compounds and minimization of derailment products are some of the factors that must be seriously considered in due time and form to attain the optimal goal of metabolic engineering.<sup>6-9</sup> On the other side, wild type genetic parts can also be evolved and optimized by different rational or random methodologies to enhance their function before the assembly or even during the integration process into the host cell. In that sense, directed evolution is an unquestionable landmark to improve enzymatic properties that cannot be found in nature as such. The generation of muteins by rounds of gene mutagenesis and the exhaustive screening of libraries have permitted the selection of useful hits. Applications and limitations of this tool box are widely reviewed and described by academic and industrial

laboratories to improve protein stability, fit the performance of new enzymes in host organisms, alter enzyme substrate specificity, and design new activities and biocatalysts.<sup>10-13</sup> Nonetheless, even if advances in directed evolution show that the principle can be applied to key components of metabolic pathways and networks, evolving complex combinations of each genetic part in a metabolic reaction cascade still remains a hard task and sometimes uncertain.<sup>5,14</sup>

Looking for a practical alternative to solve some of the technical limitations mentioned above, we have developed and recently described an efficient method that uses DNA repair deficient yeast cells capable to in vivo assemble, recombine, stably integrate, and express 23 kb plant flavonoid pathways.<sup>15</sup> In one single step of transformation this method generates functional pathway libraries containing intragenic mosaics as a result of simultaneous homeologous pairwise recombination between two wild type plant gene variants for each enzymatic reaction responsible of the synthesis of this kind of phenylpropanoids.

### Generating Mosaic Genes with DNA Repair Deficient Yeast Cells

The genetic plasticity of yeast cells makes of them a powerful tool to transfer heterologous information as genes, metabolic pathways and networks and even entire genomes by simple transformation in a targeted manner.<sup>16</sup> This information can be assembled and integrated into the chromosomes by using partially overlapping DNA fragments. The principle based on homologous recombination (recombination of identical sequences) is known as DNA Assembler.<sup>17</sup>

Homeologous recombination, defined as the recombination of DNA sequences that are not identical, is a quite uncommon event in normal cells in which several control mechanisms play the role of sentinels of the DNA integrity and the fidelity of heritable genetic information. When these mechanisms are defective in the cell, i.e., when key genes encoding repair enzymes that detect anomalies on DNA strands are deleted or inactive,

crossovers between non-identical sequences occur at high frequency in these deficient cells. For gene and pathway evolution purposes we used two deletion yeast strains:  $\Delta msh2$  and  $\Delta sgs1$ . Msh2p is a component of mismatch DNA repair complexes MutS $\alpha$  and MutS $\beta$  mainly involved in base-excision repair and homologous recombination.<sup>18,19</sup> Sgs1p belongs to the RecQ family of DNA helicases involved in genome stability and heteroduplexes rejection.<sup>19-22</sup>

Formerly we tested whether fragments containing non-identical DNA could be assembled and recombined in wild type or DNA repair deficient cells as  $\Delta msh2$ ,  $\Delta msh5$ ,  $\Delta msh6$ ,  $\Delta rad58$ , and  $\Delta sgs1$ . We therefore used different combinations of  $\beta$ -lactamase genes of the OXA class D<sup>23</sup> as recombination substrates since several candidates with different degree of DNA sequence identity (96 to 47%) were available to our purposes. DNA fragments were designed in a way that OXA ORFs serve as unique substrates for assembly and homeologous recombination (Fig. 1A). After transformation in each one of the four deletion yeast cell, the assembled clusters integrated in the target chromosome by their flanking homologous sequences and four recombinant libraries were generated. As expected, wild type cells yielded significant amount of clones only when OXA genes used in the assembly were identical or closed related (100 and 96%), confirming that DNA repair mechanisms worked in these cells avoiding the replication of large mismatched sequences. On the other hand, while  $\Delta msh5$ ,  $\Delta msh6$ , and  $\Delta rad58$  cells yielded low number of selected clones, highly rich libraries (up to 10<sup>6</sup> recombinant clones) were obtained with  $\Delta msh2$  and  $\Delta sgs1$  cells. The size of the libraries was variable in function of the DNA repair deficient cells used and the sequence similarity of genes to recombine. The number of recombinant clones generated by  $\Delta sgs1$  was about one order higher compared with those from  $\Delta msh2$ . The largest libraries were generated with OXA genes 78 to 63% related while few clones were obtained when parental genes with less of 50% of sequence identity were used.

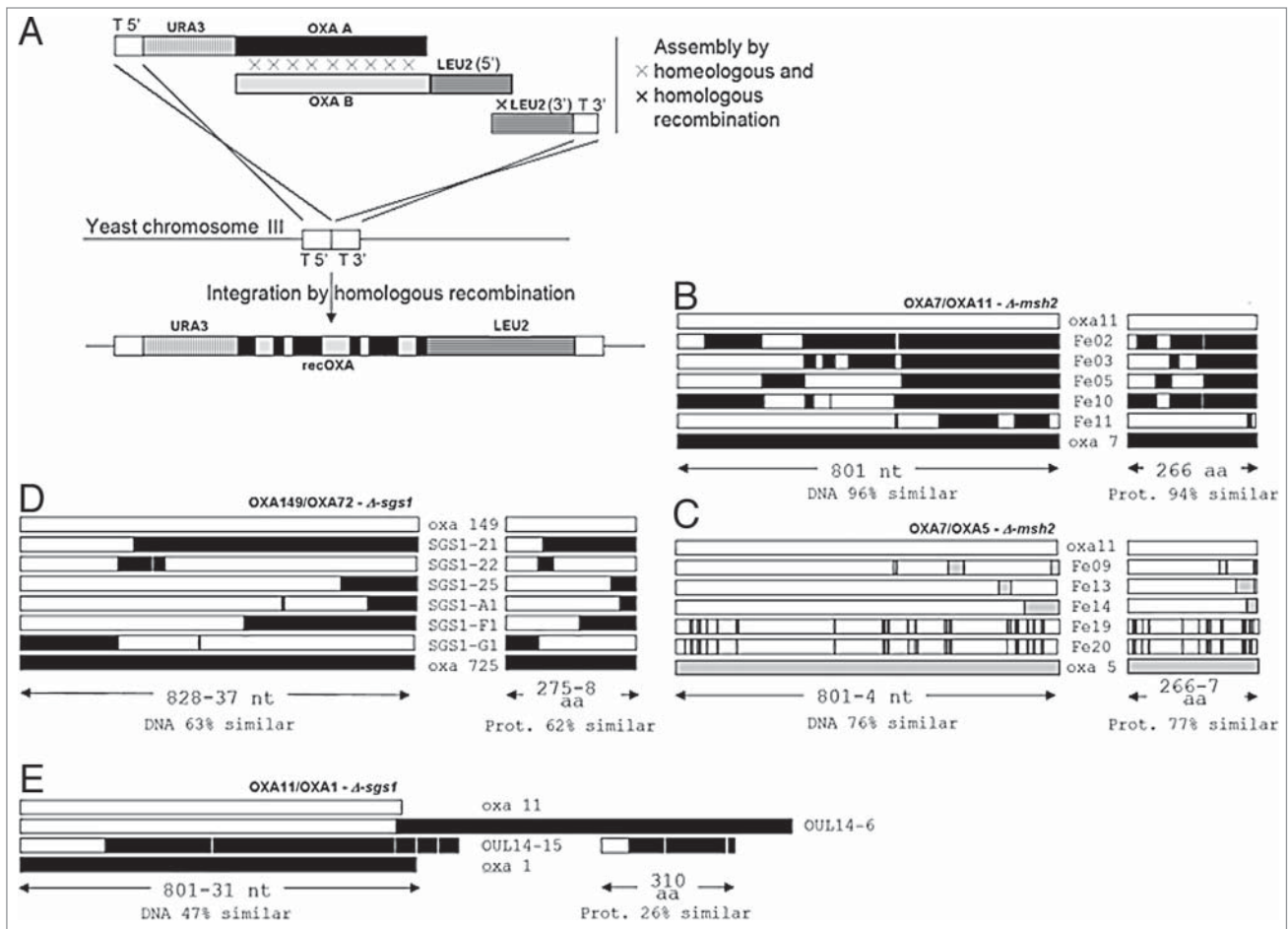
When we analyzed the recombinant OXA sequences of randomly selected clones we found different degree of mosaicism, i.e., DNA track exchanges between each pair of parental genes. We identified recombinant OXA genes with one or two exchanges as well as complex mosaics with more than 20 crossed tracks in sequences 800 base pair long. Although the parental genes had different length, each mosaic conserved an open reading frame. Surprisingly, no insertions, deletions or additional stop codons were detected. We could even observe exchanges of one single base between the two parental genes. As mosaic genes contained uninterrupted ORFs, they potentially encode mosaic proteins (Fig. 1A-E).

Since pair wise recombination of parental genes resulted in correctly assembled and recombined genes, we wanted to test if homeologous recombination with more than two parental genes was possible. Therefore, three non-identical OXA sequences (from 96 to 76% DNA related) were assembled after transformation of a  $\Delta msh2$  yeast strain (Fig. 2A). The three components generated new mosaics containing different DNA stretches from the three gene sources. Therefore, as recombinants had uninterrupted ORFs they could encode for mosaics proteins (Fig. 2B).

With this set of pilot experiments we showed that gene mosaicism by in vivo assembling and homeologous recombination in DNA repair deficient cells is a simple tool to generate rich genetic diversity.

### From Mosaic Genes to Intragenic Mosaic Pathways

These results encouraged us to assemble flavonoid gene counterparts as the sole substrates of recombination in integrative cluster units. In plants, flavonoid production starts with phenylalanine, eventually tyrosine,<sup>24</sup> as substrate which is converted by the sequential action of 8 enzymes, comprising phenylalanine ammonia lyase (PAL), cinnamate-4-hydroxylase (C4H), 4-coumaroyl-CoA ligase (4CL), chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H) and flavanol synthase (FLS), and a



**Figure 1.** Homeologous OXA genes A and B (sequence identity from 96 up to 47%) were in vivo assembled and recombined in DNA repair deficient cells  $\Delta$ -*msh2* and  $\Delta$ -*sgs1*. (A) Schematics of recombination, assembly and integration of three homeologous OXA genes in chromosome III of DNA repair deficient yeast cells. Recombination patterns of novel OXA genes in selected clones and predicted protein sequences. (B and C) White, black, and gray boxes correspond respectively to OXA11, OXA5 and OXA7 (96% and 76% similar). (D) White and black boxes correspond respectively to OXA149 and OXA72 (63% similar). (E) White and black boxes correspond respectively to OXA11 and OXA1 (47% similar). Clones were randomly isolated and sequenced after selection. Every recombinant OXA (recOXA) gene analyzed resulted in an uninterrupted open reading frame (ORF), as shown at the right of each inset.

cytochrome P450 reductase (CPR) to activate C4H. We found pairs of counterpart genes (from 96 to 75% similar at DNA level) for each enzyme of the pathway that were shown to be expressed in yeast.<sup>25-28</sup>

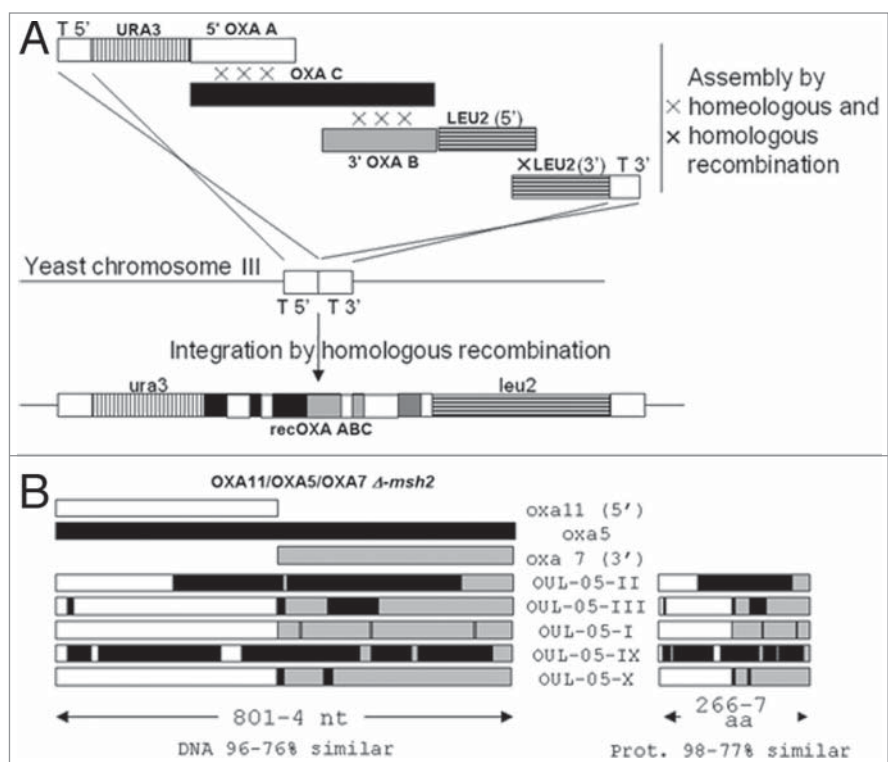
**Figure 3A** resumes the rational of our approach. DNA fragments were designed and synthesized. Combinations of these fragments containing homeologous gene versions from different plant sources were introduced in DNA repair deficient yeast strains  $\Delta$ -*msh2* and  $\Delta$ -*sgs1*. Homologous (identical) gene combinations were used as parental controls by transforming wild type yeasts. Fragments were designed in a way that once assembled and integrated the resulting recombinant pathways could

be selected (flanking auxotroph marker genes) and the mosaic genes transcribed (yeast promoters and terminators flanking each flavonoid gene). Two libraries of recombinants were generated and four clones isolated at random were analyzed.

Sequencing results showed that assembly, integration and recombination of DNA fragments were successful, rendering defined clusters of flavonoid genes stably integrated in the yeast chromosome. We found that numerous sequence exchanges defining complex mosaics occurred into the genes. As with OXAs mosaics, each flavonoid mosaic gene resulted in an uninterrupted ORF containing DNA tracks from one or the other parental source as

intragenic exchanges of variable lengths and number (Fig. 3B). Additionally, no frame shifts, deletions, or insertions in generated mosaic ORFs were detected. Thus, we proved that our technology is a simple and useful method to generate libraries containing distinct and novel intragenic mosaics pathways.

These recombined mosaic pathways could only be observed in DNA repair deficient cells. Although in vivo homeologous recombination has been described in wild type cells,<sup>29-32</sup> there seem to be a limit in the number and size of fragments and their nucleotide divergence preventing the correct recombination of non-identical sequences in DNA repair proficient cells. It was reported that double and single



**Figure 2.** Homeologous OXA genes A, B, and C (sequence identity from 96 up to 76%) were *in vivo* assembled and recombined in DNA repair deficient cells  $\Delta$ -*msh2* and  $\Delta$ -*sgs1*. **(A)** Schematics of recombination, assembly and integration of three homeologous OXA genes in chromosome III of DNA repair deficient yeast cells. Recombination patterns of novel OXA genes in selected clones and predicted protein sequences. **(B)** White, black, and gray boxes correspond respectively to OXA11, OXA and OXA7 (from 96% to 76% similar). Clones were randomly isolated and sequenced after selection. Every recOXA gene analyzed resulted in an open reading frame (ORF), as shown at the right of the inset.

strands breaks license chromatid exchanges.<sup>19-21,33</sup> The input of synthetic fragments in our method actually represents an important source of free recombinogenic double strand ends in the cell that could mimic damaged DNA. In the absence of Msh2p or Sgs1p, proofreading and discrimination between identical and non-identical DNA are perturbed, although the rest of repair mechanisms of the cell remains functional. Recently it has been shown that Holliday junctions in a gap containing plasmid which was used to transform a  $\Delta$ -*sgs1* yeast strain cannot be resolved because this cell failed in repairing and rejecting heteroduplexes on the annealed structure formed (D-loop) between the plasmid and the genome.<sup>34</sup> Thus, this  $\Delta$ -*sgs1* strain incorporates mismatched tracks of DNA into the cell genome or plasmid, which results in mosaic sequences of different extents.

Hence, DNA repair deficient cells can efficiently assemble homeologous fragments to generate mosaic genes and intragenic mosaic pathways. We cannot conclude by our sequence analysis if different repair deficient mutants produced different types of mosaics. Nonetheless  $\Delta$ -*sgs1* strain generated a higher number of unique recombinants using the same transformation protocol than  $\Delta$ -*msh2*, which suggest that  $\Delta$ -*sgs1* background is more appropriate to generate large libraries.

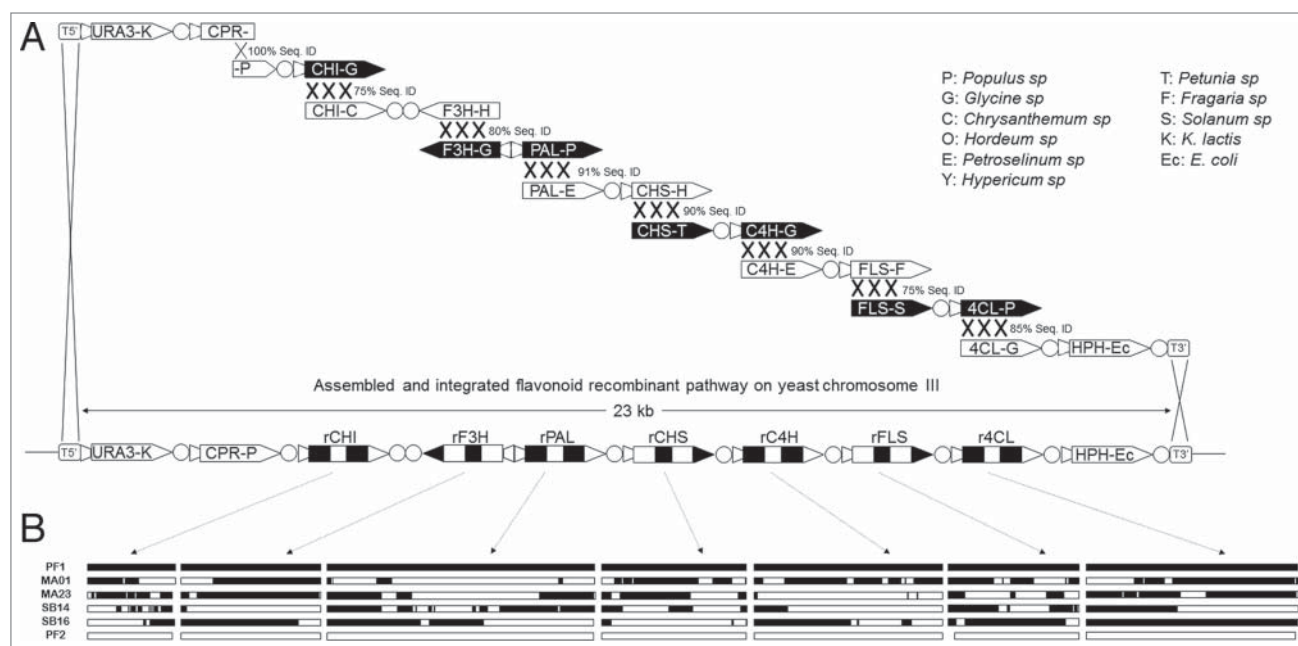
DNA repair deficient cells are therefore a tool to generated mosaicism from counterpart genes that are not licensed to recombine in wild type cells. Nonetheless, this deficiency condition must be controlled and avoided once mosaic clones are obtained. Since *in vivo* evolution is performed in haploids cells, we implemented the sexual conjugation ability of budding yeast to generate diploids: as

soon as haploid recombinant clones are selected, they are crossed with a wild type mating haploid (i.e., DNA repair proficient *Msh2* or *Sgs1*) which permits the restoration of all DNA repair functions in the diploid cell newly generated as well as the maintaining of their new metabolic functions previously obtained. In that way, cells become genetically stable and could be used as such for production purposes. Moreover, as we generated mosaic haploids and restore the DNA repair proficiency (the absence of Msh2p in one haploid is complemented by the other which contains the gene; similarly, the deletion of *Sgs1* in the later is complemented by the active gene in the former). Thus, this diploid cell becomes genetically stable and now contains two “loads” of mosaic pathways which shows, as expected, an increased flavonoid production level (see Fig. 4, inset C in our article<sup>15</sup>).

## Functionality of Intragenic Pathways and Metabolic Diversity

When supernatants of selected parental and intragenic mosaics clones cultured in synthetic growth media were analyzed by HPLC and LC/MS/MS we found and characterized several flavonoids and intermediate metabolites present at different extent in parental clones and recombinant mosaics. Figure 4 is a good example of differences in production profiles of parental and mosaic clones. In fact, these results represented the first evidence that intragenic mosaicism of recombinant pathways generated by *in vivo* recombination generates fully functional and diverse pathways. Intragenic mosaic clones presented a production profile different to those observed in parental controls, meaning that functionality of recombinants also reflects different degrees of metabolic diversity. We found that one of the analyzed clones containing an intragenic mosaic pathway (SB16) synthesized large amounts of naringenin, dihydrokaempferol, and kaempferol, compared with both parental





**Figure 3.** In vivo assembly, homeologous recombination and chromosomal integration of flavonoid intragenic mosaic pathways. **(A)** Scheme of 9 synthetic fragments used for transformation of DNA repair deficient yeast strains. In one single step, fragments were assembled in a pairwise manner with homeologous genes as sole substrates of recombination. Sequence identities of the genes are indicated as% of matching nucleotides for each ORF. X represents recombination events. Fragments 1 and 9 contain the 5' and 3' integration targets in loci bud31-hcm1 of the yeast chromosome III, as well as the auxotroph marker URA3 and the gene encoding the resistance to hygromycin (HPH) used for selection of recombinants. Each homeologous versions for the 7 flavonoid genes are schematized in black and white boxes in alternating fragments 2 to 8. Plant sources of the genes are indicated. **(B)** Mosaicism of homeologous recombinant genes after transformation into DNA repair deficient cells. Mosaic sequences detected after sequencing of randomly isolated clones (MA01, MA23, SB14, and SB16) are represented as successive white and black tracks. Parental genes PF1 and PF2 (black and white boxes) generated by homologous recombination contain 100% of nucleotide sequence identity respectively. Adapted from Luque et al.<sup>15</sup>

controls. Other clone (MA01) mainly accumulated the early flavonoid intermediate p-coumaric acid at high concentrations.

Several other peaks were detected in culture supernatant of intragenic mosaic pathway clones as phloretic acid and phloretin, and aromatic derivatives such as styrene and p-hydroxystyrene. These last two molecules were detected in all recombinant clones suggesting that they are derivatives of the main flavonoid metabolic flux. This derailment is certainly due to the action of endogenous yeast decarboxylases using early products of the flavonoid metabolism, mainly the cinnamic acid.<sup>35,36</sup> As for phloretic acid, it is likely that synthesized p-coumaroyl-CoA is transformed by endogenous yeast dehydrogenases for the conversion into p-dihydrocoumaroyl-CoA and further to phloretic acid.<sup>37</sup> Other eluted peaks were also detected only in recombinant clones but could not be elucidated by our means.

However, these compounds were certainly derivatives of the flavonoid gene pathways since they are undetectable in wild type cells.

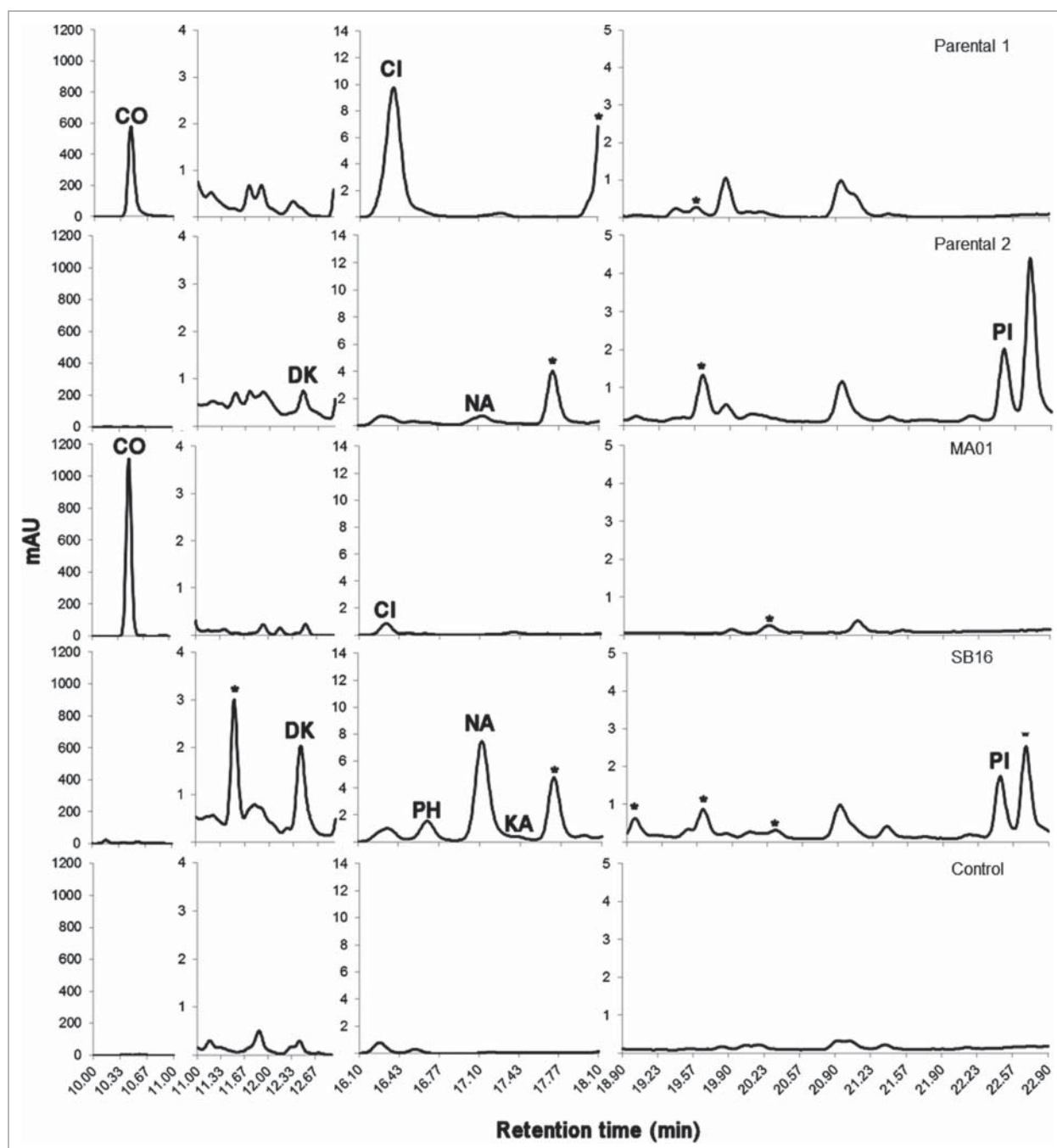
These results strengthen the concept that the recombined pathways are functional and that a complex phenylpropanoid compound library has been generated by our method. Moreover, the different expression of certain mosaic enzymes generated a landscape of diverse flavonoid metabolic profiles conditioned by differences in the production rate of each mosaic pathway.

### Induced Expression of Intragenic Pathways to Analyze Enzymatic Steps

The fact that the total production of phenylpropanoids from intragenic mosaic pathways was different to that of parental ones suggested that production

rates and enzymatic reactions behave distinctly in different recombinants. The rationale of our approach permitted to explore the ability of mosaic clones to convert, as an example, cinnamic acid, naringenin, and dihydrokaempferol when added as precursors in partial induction media. In such media, the addition or suppression of certain compounds either repress or induce the transcription of certain genes, i.e., *Gal*, *Met*, and *Trp* promoters that were added to some genes of the pathway. This enabled a better control of parental and mosaic enzymes PAL, C4H, CHI, F3H, and FLS expressed in the context of a partially functional metabolic pathway.

By these means we showed that mosaic FLS enzyme from one recombinant was 5 times more efficient in converting dihydrokaempferol to final flavonoid product Kaempferol compared with each one of its wild type counterparts. We also found that another recombinant used the feeder

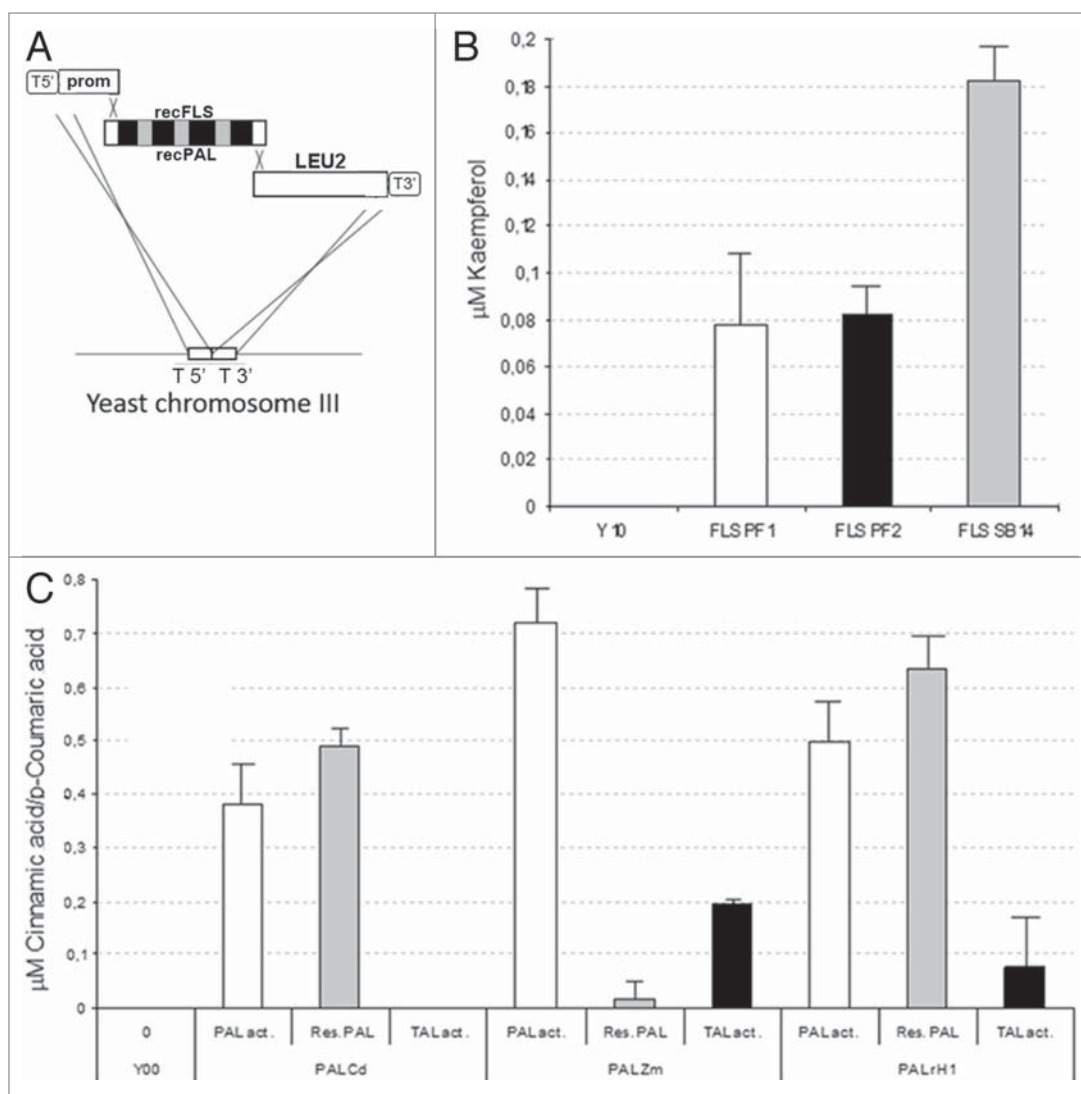


**Figure 4.** Examples of functionality of mosaic pathways as detected by HPLC and LC/MS/MS. UV/Vis chromatogram of culture supernatants detected at 290 nm of parental clones 1 and 2 and two intragenic mosaic clones (MA01, and SB16). Negative control corresponds to the untransformed yeast cell. CO, p-coumaric acid; CI, cinnamic acid; NA, naringenin; DK, dihydrokaempferol; K, kaempferol; PI, pinocembrin; PH, phloretin. Non-identified (flavonoid specific) peaks are labeled with a star (\*). Histograms were fractionated and mAU scale adapted to each fraction for a better visualization of revealed peaks.

cinnamic acid twice better than other clones, which could explain the accumulation of p-coumaric acid observed in this clone in fully induction conditions. These results suggest that even if individual mosaic genes encode more efficient enzymes compared with parental ones,

clones will tend to accumulate metabolic intermediates or final products depending on the enzymatic context of their entire and fully functional mosaic pathway. In other words, gene mosaicism can influence metabolic fluxes and thereby production rates.

By the way, several bottlenecks have been observed in the production rate of flavonoids and the availability of certain precursors. Malonyl-CoA seems to have a major influence in the balance of flavonoid precursors and final products.<sup>38,39</sup> We observed traces of flavanones in



**Figure 5.** Genes isolated from intragenic mosaic pathways and their enzymatic diversity as a result of in vivo homeologous recombination. **(A)** Mosaic FLS and PAL genes were isolated from selected clones and re-integrated alone out of the context of the flavonoid pathway in wild type cells by using a classical integration approach. Used promoters (prom) were pADH1 for FLS and pGAL10 for PAL **(B)** Mosaic FLS gene from one of the clones analyzed (SB14) encodes an enzyme (grey box) with an increased ability to convert the substrate dihydrokaempferol in the product kaempferol when compared with wild type parental genes (white and black boxes, PF1 and PF2 respectively). **(C)** Mosaic PAL (rh1) is a new PAL enzyme containing both, PAL heat resistant (grey box) and TAL activity (black box) inherited from *Zea mays* (Zm) and *Cistanche deserticola* (Cd) related genes after homeologous recombination. Individual clones were cultured by triplicate and standard deviations are shown over the bars.

supernatants of one of the parental clones. This could be the result of the *Glycine max* 4CL enzyme which activity is not enough to process the activated malonate available in the yeast host. Metabolic engineering modifications in yeast cells increasing the amounts of this precursor (i.e., by coupling the overexpression of malonate synthase genes and the downregulation of fatty-acid metabolism) as recently shown in *E. coli*<sup>40</sup> should be considered for the optimization of flavonoids in yeast cell factories.

### In Vivo Evolution in DNA Repair Deficient Cells and Generation of Novel Functions

We wondered if different mosaic clone behaviors in the flavonoid pathway context could be correlated at the level of single mosaic enzymes. For that we isolated the genes, both parental and mosaic, and expressed them out in wild type yeasts. To do so, FLS genes were amplified from gDNA of parent and mosaic clones and then integrated in a

wild type yeast strain (Fig. 5A). Selected clones were cultured synchronously by triplicates in media containing the substrate dihydrokaempferol. Supernatants were recovered at different time points (up to 7 h) and analyzed by HPLC. Average amounts of produced kaempferol were determined for each clone (Fig. 5B).

Mosaic and parental FLS enzymes synthesized kaempferol to different extents when isolated genes were expressed in the cells. While the parental enzymes (black

and white squares) produced basically the same amount of the metabolite, the mosaic FLS of one clone (gray square) showed a considerable increased production of kaempferol, twice more than parental producers. This behavior is also consistent with the fact that this clone was the best producer of kaempferol in the mosaic pathway context when the clone was cultured in partially induced conditions and fed with dihydrokaempferol. FLS expression results suggest that even if mosaic enzyme of this recombinant is able to convert more dihydrokaempferol in kaempferol than parental ones, the production rate of the pathway depends on the availability of earlier metabolites (naringenin and dihydrokaempferol) supplied by former enzymes CHI and F3H. In other words, we generated a mosaic FLS gene encoding a functional enzyme conferring to the cell the ability to convert more substrate than wild type counterpart genes when the compound is available from earlier reactions of the pathway.

The mosaicism obtained by this *in vivo* method not only affects enzymatic activities of recombinant proteins it generates, but also can deeply modify their species specific enzymatic properties in a very attractive way. As already mentioned, PAL is the first enzyme of flavonoid pathway that transforms the input L-phenylalanine (Phe) in cinnamic acid or L-tyrosine (Tyr) in p-coumaric acid, both metabolites primarily essential for the phenylpropanoid metabolic flow. It is well known that dicotyledonous PALs are strictly PAL enzymes (i.e., TAL activity, the ability to use Tyr, is almost inexistent). On the other hand, monocotyledonous PALs have both PAL and TAL activities. Thus, monocotyledonous *Zea mays* (Zm) PAL can take both Phe and Tyr as substrates,<sup>41</sup> whereas the dicotyledonous *Cistanche deserticola* (Cd) PAL have a strict PAL activity. Additionally Cd PAL is heat shock resistant: the protein retains 87% of its activity after heating (20 minutes at 70 °C).<sup>42</sup>

When both genes were subjected to *in vivo* evolution in DNA repair deficient cells, libraries were generated and clones were screened in their sensibility and/or resistance after heating and to use both Phe and Tyr as substrates. Proceeding in

the same manner as with FLS recombinants, amid the PAL/TAL clones analyzed one of them with a single crossover showed a strong resistance to high temperatures and both, TAL and PAL activities were retained (Fig. 5C). This result, that will be part of a next publication, enforces the potential of *in vivo* evolution to generate from old parts new interesting diversity and functions that are not found in nature.

Dicotyledonous encode the TAR activity in several genes, meaning that they can also drive tyrosine to the phenylpropanoid network by this way. PAL genes are strongly well conserved in evolution and their basic functions seem to predate the divergence monocotyledonous/dicotyledonous. There exists some evidence concerning the specialization of this gene family after the divergence, which could explain why monocots conserved the TAR function in PAL genes unlike dicots.<sup>43</sup> It must be kept in mind that most current monocots have been suffered a strong human selection by their agricultural interest and therefore the specificity of genes intervening in the phenylpropanoid metabolism could have been modified. This could be also the case with *Cistanche deserticola*,<sup>42</sup> a dicotyledonous plant very praised by its natural medicinal properties in Asia. PAL enzyme (and maybe others) shows a marked heat resistance, an interesting trait that could have been selected and fixed as consequence of its wide consumption by local populations.

## Concluding Remarks

An efficient and simple *in vivo* evolution strategy to assemble, recombine and express complex libraries of a plant metabolic pathway has been recently described.<sup>15</sup> Synthetic overlapping and non-identical DNA fragments containing counterpart genes of different origins as sole substrates for assembly and recombination were designed and used to transform DNA repair deficient cells. The method generates complex mosaic genes from parental sequences as divergent as 50% at DNA level. For metabolic pathway assembly and evolution, 7 pairs of counterpart flavonoid genes from 91% up

to 75% similar at DNA level were used to successfully generate libraries of 23 kb intragenic mosaic pathways. An eighth gene (CPR) was also added to the cells for the activation of C4H enzyme. Despite these gene divergences, our *in vivo* recombination system produced mosaic open reading frames conserving original gene structures that were correctly located between their designated promoters and terminators and stably integrated in the target yeast chromosome.

We have shown that our method enabled the rapid evolution of metabolic pathways resulting in the production of different final metabolites. In our model, the production ratio of recombinants was conditioned by the molecular diversity generated from intragenic mosaic pathways. The recovery of entire mosaic pathways of interest is feasible; therefore our strategy used in an iterative way (i.e., by recovering interesting hits and used as new substrates of homeologous recombination) becomes a reasonable possibility that should include the *in vivo* homeologous recombination and assembly in DNA repair deficient cells as an additional evolution approach for the synthetic biology toolbox. This technology can also be applied for the rapid evolution of other pathways implementing complementary approaches in which optimized host cells become metabolic chassis to improve the production of precursors and to obtain large amounts of end products and metabolic intermediates. The metabolic derailment is often an insidious problem because it reflects that the highly adaptable metabolic machinery of the cell is working out of the bioengineer's control. That is why a cell that optimally expresses heterologous genes is not necessarily enough to optimally produce the metabolite of interest if some intermediates are used by the host. These problems could be solved by modifying the causes of the derailment when they are known or suspected, and if they are not critical or essential for cell viability, as recently shown for the flavonoid naringenin.<sup>44</sup> In these cells, the inactivation of genes controlling the synthesis of aromatic amino acids (*Aro3* and *Aro4*) alleviated the feedback inhibition of phenylalanine and tyrosine which optimized their participation in the flavonoid flow.



In a same way, the undesirable derivation of p-coumaric acid by host decarboxylases was avoided by deleting *Aro10*, *Pdc1*, *Pdc5*, and *Pdc6* host genes.

Since DNA repair mechanisms are highly conserved between the different species, is reasonable to think that this technology could be applied to recombine metabolic pathways in other production organisms. Nevertheless, bioengineering hints discussed throughout this commentary should be considered in each case. Taken together, our approach offers an exciting alternative to molecular engineering, combinatorial and semi-rational directed evolution methods aiming for increasing the production, or the generation of new compounds and to explore metabolic fluxes in a defined cellular context by controlling the expression of evolved pathway components.

#### Disclosure of Potential Conflicts of Interest

No potential conflict of interest was disclosed.

#### References

- Stephanopoulos G. Synthetic biology and metabolic engineering. *ACS Synth Biol* 2012; 1:514-25; PMID:23656228; <http://dx.doi.org/10.1021/sb300094q>
- König H, Frank D, Heil R, Coenen C. Synthetic genomics and synthetic biology applications between hopes and concerns. *Curr Genomics* 2013; 14:11-24; PMID:23997647
- Cameron DE, Bashor CJ, Collins JJ. A brief history of synthetic biology. *Nat Rev Microbiol* 2014; 12:381-90; PMID:24686414; <http://dx.doi.org/10.1038/nrmicro3239>
- Khalil AS, Collins JJ. Synthetic biology: applications come of age. *Nat Rev Genet* 2010; 11:367-79; PMID:20395970; <http://dx.doi.org/10.1038/nrg2775>
- Quin MB, Schmidt-Dannert C. Engineering of biocatalysts - from evolution to creation. *ACS Catal* 2011; 1:1017-21; PMID:22125758; <http://dx.doi.org/10.1021/cs200217t>
- Umenhoffer K, Fehér T, Balikó G, Ayaydin F, Pósfai J, Blattner FR, Pósfai G. Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microb Cell Fact* 2010; 9:38; PMID:20492662; <http://dx.doi.org/10.1186/1475-2859-9-38>
- Danchin A. Scaling up synthetic biology: Do not forget the chassis. *FEBS Lett* 2012; 586:2129-37; PMID:22226636; <http://dx.doi.org/10.1016/j.febslet.2011.12.024>
- Carbonell P, Fichera D, Pandit SB, Faulon JL. Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Syst Biol* 2012; 6:10; PMID:22309974; <http://dx.doi.org/10.1186/1752-0509-6-10>
- Kim IK, Roldão A, Siewers V, Nielsen J. A systems-level approach for metabolic engineering of yeast cell factories. *FEMS Yeast Res* 2012; 12:228-48; PMID:22188344; <http://dx.doi.org/10.1111/j.1567-1364.2011.00779.x>
- Eijsink VGH, Gåseidnes S, Borchert TV, van den Burg B. Directed evolution of enzyme stability. *Biomol Eng* 2005; 22:21-30; PMID:15857780; <http://dx.doi.org/10.1016/j.bioeng.2004.12.003>
- Parales RE, Ditty JL. Laboratory evolution of catabolic enzymes and pathways. *Curr Opin Biotechnol* 2005; 16:315-25; PMID:15961033; <http://dx.doi.org/10.1016/j.copbio.2005.03.008>
- Dougherty MJ, Arnold FH. Directed evolution: new parts and optimized function. *Curr Opin Biotechnol* 2009; 20:486-91; PMID:19720520; <http://dx.doi.org/10.1016/j.copbio.2009.08.005>
- Urvoas A, Valerio-Lepiniec M, Minard P. Artificial proteins from combinatorial approaches. *Trends Biotechnol* 2012; 30:512-20; PMID:22795485; <http://dx.doi.org/10.1016/j.tibtech.2012.06.001>
- Cobb RE, Sun N, Zhao H. Directed evolution as a powerful synthetic biology tool. *Methods* 2013; 60:81-90; PMID:22465795; <http://dx.doi.org/10.1016/j.ymeth.2012.03.009>
- Luque A, Sebai SC, Santiago-Schübel B, Le Coz Y, Jenot D, Ramaen O, Sauveplane V, Pandjaitan R. In vivo evolution of metabolic pathways by homologous recombination in mitotic cells. *Metab Eng* 2014; 23:123-35; PMID:24685654; <http://dx.doi.org/10.1016/j.ymben.2014.02.010>
- Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A* 2008; 105:20404-9; PMID:19073939; <http://dx.doi.org/10.1073/pnas.0811011106>
- Shao Z, Zhao H, Zhao H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* 2009; 37:e16; PMID:19074487; <http://dx.doi.org/10.1093/nar/gkn991>
- Junop MS, Yang W, Funchain P, Clendenin W, Miller JH. In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair (Amst)* 2003; 2:387-405; PMID:12606120; [http://dx.doi.org/10.1016/S1568-7864\(02\)00245-8](http://dx.doi.org/10.1016/S1568-7864(02)00245-8)
- Hoffmann ER, Borts RH. Trans events associated with crossovers are revealed in the absence of mismatch repair genes in *Saccharomyces cerevisiae*. *Genetics* 2005; 169:1305-10; PMID:15654113; <http://dx.doi.org/10.1534/genetics.104.033407>
- Li GM. Mechanisms and functions of DNA mismatch repair. *Cell Res* 2008; 18:85-98; PMID:18157157; <http://dx.doi.org/10.1038/cr.2007.115>
- Schmidt KH, Wu J, Kolodner RD. Control of translocations between highly diverged genes by Sgs1, the *Saccharomyces cerevisiae* homolog of the Bloom's syndrome protein. *Mol Cell Biol* 2006; 26:5406-20; PMID:16809776; <http://dx.doi.org/10.1128/MCB.00161-06>
- Amin AD, Chaix ABH, Mason RP, Badge RM, Borts RH. The roles of the *Saccharomyces cerevisiae* RecQ helicase Sgs1 in meiotic genome surveillance. *PLoS One* 2010; 5:e15380; PMID:21085703; <http://dx.doi.org/10.1371/journal.pone.0015380>
- Bush K, Jacoby GA. Updated functional classification of beta-lactamases. *Antimicrob Agents Chemother* 2010; 54:969-76; PMID:19995920; <http://dx.doi.org/10.1128/AAC.01009-09>
- Buer CS, Imin N, Djordjevic MA. Flavonoids: new roles for old molecules. *J Integr Plant Biol* 2010; 52:98-111; PMID:20074144; <http://dx.doi.org/10.1111/j.1744-7909.2010.00905.x>
- Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 2001; 126:485-93; PMID:11402179; <http://dx.doi.org/10.1104/pp.126.2.485>
- Trantas E, Panopoulos N, Ververidis F. Metabolic engineering of the complete pathway leading to heterologous biosynthesis of various flavonoids and stilbenoids in *Saccharomyces cerevisiae*. *Metab Eng* 2009; 11:355-66; PMID:19631278; <http://dx.doi.org/10.1016/j.ymben.2009.07.004>
- Naesby M, Nielsen SV, Nielsen CA, Green T, Tange TO, Simón E, Knechtel P, Hansson A, Schwab MS, Titiz O, et al. Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in *Saccharomyces cerevisiae*. *Microb Cell Fact* 2009; 8:45; PMID:19678954; <http://dx.doi.org/10.1186/1475-2859-8-45>
- Santos CNS, Koffas M, Stephanopoulos G. Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 2011; 13:392-400; PMID:21320631; <http://dx.doi.org/10.1016/j.ymben.2011.02.002>
- Rayssiguier C, Thaler DS, Radman M. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* 1989; 342:396-401; PMID:2555716; <http://dx.doi.org/10.1038/342396a0>
- Datta A, Adjiri A, New L, Crouse GF, Jinks Robertson S. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. *Mol Cell Biol* 1996; 16:1085-93; PMID:8622653
- Elefany AG, Begley CG, Metcalf D, Barnett L, Köntgen F, Robb L. Characterization of hematopoietic progenitor cells that express the transcription factor SCL, using a lacZ "knock-in" strategy. *Proc Natl Acad Sci U S A* 1998; 95:11897-902; PMID:9751762; <http://dx.doi.org/10.1073/pnas.95.20.11897>
- Swers JS, Kellogg BA, Wittrop KD. Shuffled antibody libraries created by *in vivo* homologous recombination and yeast surface display. *Nucleic Acids Res* 2004; 32:e36; PMID:14978223; <http://dx.doi.org/10.1093/nar/gnh030>
- Helleday T, Lo J, van Gent DC, Engelward BP. DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)* 2007; 6:923-35; PMID:17363343; <http://dx.doi.org/10.1016/j.dnarep.2007.02.006>
- Mitchel K, Lehner K, Jinks-Robertson S. Heteroduplex DNA position defines the roles of the Sgs1, Srs2, and Mph1 helicases in promoting distinct recombination outcomes. *PLoS Genet* 2013; 9:e1003340; PMID:23516370; <http://dx.doi.org/10.1371/journal.pgen.1003340>
- Takemoto M, Achiwa K. Synthesis of styrenes through the biocatalytic decarboxylation of trans-cinnamic acids by plant cell cultures. *Chem Pharm Bull (Tokyo)* 2001; 49:639-41; PMID:11383623; <http://dx.doi.org/10.1248/cpb.49.639>
- Ro DK, Douglas CJ. Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): implications for control of metabolic flux into the phenylpropanoid pathway. *J Biol Chem* 2004; 279:2600-7; PMID:14607837; <http://dx.doi.org/10.1074/jbc.M309951200>
- Gosch C, Halbwirth H, Stich K. Phloridzin: biosynthesis, distribution and physiological relevance in plants. *Phytochemistry* 2010; 71:838-43; PMID:20356611; <http://dx.doi.org/10.1016/j.phytochem.2010.03.003>
- Zha W, Rubin-Pitel SB, Shao Z, Zhao H. Improving cellular malonyl-CoA level in *Escherichia coli* via metabolic engineering. *Metab Eng* 2009; 11:192-8; PMID:19558964; <http://dx.doi.org/10.1016/j.ymben.2009.01.005>
- Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MA. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 2011; 13:578-87; PMID:21763447; <http://dx.doi.org/10.1016/j.ymben.2011.06.008>
- Wu J, Du G, Zhou J, Chen J. Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab Eng* 2013; 16:48-55; PMID:23246524; <http://dx.doi.org/10.1016/j.ymben.2012.11.009>

41. Rösler J, Krekel F, Amrhein N, Schmid J. Maize phenylalanine ammonia-lyase has tyrosine ammonia-lyase activity. *Plant Physiol* 1997; 113:175-9; PMID: 9008393; <http://dx.doi.org/10.1104/pp.113.1.175>
42. Hu GS, Jia JM, Hur YJ, Chung YS, Lee JH, Yun DJ, Chung WS, Yi GH, Kim TH, Kim DH. Molecular characterization of phenylalanine ammonia lyase gene from *Cistanche deserticola*. *Mol Biol Rep* 2011; 38:3741-50; PMID:21104014; <http://dx.doi.org/10.1007/s11033-010-0489-0>
43. Rawal HC, Singh NK, Sharma TR. Conservation, Divergence, and Genome-Wide Distribution of PAL and POX A Gene Families in Plants. *Int J Genomics* 2013; 2013:678969; PMID:23671845; <http://dx.doi.org/10.1155/2013/678969>
44. Koopman F, Beekwilder J, Crimi B, van Houwelingen A, Hall RD, Bosch D, van Maris AJ, Pronk JT, Daran J-M. De novo production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. *Microb Cell Fact* 2012; 11:155-70; PMID:23216753; <http://dx.doi.org/10.1186/1475-2859-11-155>