



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 5/04, 15/29, 15/82, A01H 4/00	A1	(11) International Publication Number: WO 00/12680 (43) International Publication Date: 9 March 2000 (09.03.00)
(21) International Application Number: PCT/US99/19732 (22) International Filing Date: 30 August 1999 (30.08.99) (30) Priority Data: 60/098,528 31 August 1998 (31.08.98) US (71) Applicant (for all designated States except US): MAXYGEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95032 (US). SUBRAMANIAN, Venkiteswaran [US/US]; 3980 Corte Mar de Hierba, San Diego, CA 92130 (US). (74) Agents: QUINE, Jonathan, Alan; Law Offices of Jonathan Alan Quine, P.O. Box 458, Alameda, CA 94501 (US) et al.		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: TRANSFORMATION, SELECTION, AND SCREENING OF SEQUENCE-SHUFFLED POLYNUCLEOTIDES FOR DEVELOPMENT AND OPTIMIZATION OF PLANT PHENOTYPES (57) Abstract The invention relates to methods and compositions for generating, modifying, adapting, and optimizing polynucleotide sequences that confer detectable phenotypic properties on plant species, and related aspects.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5 **TRANSFORMATION, SELECTION, AND SCREENING OF SEQUENCE-
SHUFFLED POLYNUCLEOTIDES FOR DEVELOPMENT AND
OPTIMIZATION OF PLANT PHENOTYPES**

CROSS REFERENCE TO RELATED APPLICATIONS

 This application is a non-provisional filing of provisional application
USSN 60/098,528, filed August 31, 1998, entitled "TRANSFORMATION,
10 SELECTION, AND SCREENING OF SEQUENCE-SHUFFLED
POLYNUCLEOTIDES FOR DEVELOPMENT AND OPTIMIZATION OF PLANT
PHENOTYPES" by Willem P.C. Stemmer and Venkiteswaran Subramanian.

FIELD OF THE INVENTION

 The invention relates to methods and compositions for generating,
15 modifying, adapting, and optimizing polynucleotide sequences that confer detectable
phenotypic properties on plant species, agronomically-important microorganisms,
genetic constructs/vectors, and related aspects.

BACKGROUND

GENETIC ENGINEERING OF AGRICULTURAL ORGANISMS

20 Genetic engineering of agricultural organisms dates back thousands of
years to the dawn of agriculture. Agricultural organisms having phenotypic traits that
were deemed desirable have been selected, including taste, high yield, caloric value,
ease of propagation, resistance to pests and disease, and appearance. Classical
breeding methods to select for germplasm encoding desirable agricultural traits had
25 been a standard practice of the world's farmers long before Gregor Mendel and others
identified the basic rules of segregation and selection. For the most part, the
fundamental process underlying the generation and selection of desired traits was the
natural mutation frequency and recombination rates of the organisms, which are quite
slow compared to the human lifespan and make it difficult to use conventional
30 methods of breeding to rapidly obtain or optimize desired traits in an organism.

 The very recent advent of non-classical, or "recombinant" genetic
engineering techniques has provided a new approach for generating agricultural

organisms having desired traits and providing an economic, ecological, nutritional, or aesthetic benefit. To date, most recombinant approaches have involved transferring a novel or modified gene into the germline of an organism to effect its expression or to inhibit the expression of the endogenous homologue gene in the organism's native genome. However, the currently used recombinant techniques are generally unsuited for substantially increasing the rate at which a novel or improved phenotypic trait can be evolved. Essentially all recombinant genes in use today for agriculture are obtained from the germplasm of existing plant and microbial specimens, which have naturally evolved coordinately with constraints related to other aspects of the organism's evolution and typically are not optimized for the desired phenotype(s). The sequence diversity available is limited by the natural genetic variability within the existing specimen gene pool, although crude mutagenic approaches have been used to add to the natural variability in the gene pool.

Unfortunately, the induction of mutations to generate diversity often requires chemical mutagenesis, radiation mutagenesis, tissue culture techniques, or mutagenic genetic stocks. These methods provide means for increasing genetic variability in the desired genes, but frequently produce deleterious mutations in many other genes. These other traits may be removed, in some instances, by further genetic manipulation (e.g., backcrossing), but such work is generally both expensive and time consuming. For example, in the flower business, the properties of stem strength and length, disease resistance and maintaining quality are important, but are often initially compromised in a mutagenesis process.

As noted, the advent of recombinant DNA technology has provided agriculturists with additional means of modifying plant genomes. While certainly practical in some areas, to date, genetic engineering methods have had limited success in transferring or modifying important biosynthetic or other pathways.

Thus, there exists a need for improved methods for producing plants and agricultural microbes with desired phenotypic traits. In particular, these methods should provide general ways for achieving phenotypic modification, including increasing the diversity of the gene pool and the rate at which genetic sequences encoding desired traits are evolved, and may lessen or eliminate entirely the necessity for performing expensive and time-consuming conventional breeding and backcrossing. It is particularly desirable to have methods which are suitable for rapid

evolution of genetic sequences to function in one or more plant species and confer a desired phenotype to plants which express the genetic sequence(s).

The present invention meets these and other needs and provides such improvements and opportunities.

5 The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. All publications cited are incorporated herein by reference, whether specifically noted as such or not.

10 SUMMARY OF THE INVENTION

 The present invention provides a composition comprising a population of protoplast library members, wherein said protoplast library members each comprise a plant cell protoplast harboring intracellularly one or a subset of a library of heterologous polynucleotide sequences, each of which is operably linked to an
15 expression sequence, or, if the heterologous polynucleotide sequence is a transcriptional regulatory sequence, operably linked to a reporter gene sequence. The library of heterologous polynucleotide sequences comprise at least 10, usually at least 100, and typically at least 1,000 species of distinct heterologous polynucleotide sequences which, in certain embodiments, may share 70 to 99 percent sequence
20 identity or more, and/or may differ by only one or several nucleotide differences, and/or may share less than 70 sequence identity, or a combination thereof. Typically, the heterologous polynucleotide sequence is xenogenic; however in some embodiments the heterologous polynucleotides may be derived from genetic sequences from the genome of the same plant species from which the plant cell
25 protoplast was produced, but said heterologous polynucleotides are not naturally-occurring sequences in said genome and comprise at least one mutation or recombination not present in the genome of the plant cell protoplast.

 Most usually the heterologous sequence is substantially identical to a naturally-occurring gene sequence in the genome of a species of plant, algae,
30 dinoflagellate, bacterium, archaebacterium, cyanobacterium, plant pathogen (insect, nematode, virus, fungus), which is substantially or completely absent in the genome of the plant species from which the plant cell protoplast was produced. In an aspect, the protoplast library members comprise an expression library of cloned heterologous

polynucleotides, such as an expression cDNA library, transformed by suitable means into said plant cell protoplasts. In an aspect, the protoplast library members contain heterologous polynucleotides which are sequence-shuffled variants of at least two parental polynucleotide species, which typically share at least 70 percent sequence identity or which contain site-specific recombination sequences, or compatible restriction sites which can be used for cassette shuffling, or a combination thereof. In an embodiment, the invention provides a plant cell protoplast library comprising a plurality of library members, wherein each library member comprises a plant protoplast containing an intracellular polynucleotide comprising a distinct species of heterologous polynucleotide sequence operably linked to an expression sequence (e.g., a transcriptional regulatory sequence functional in the protoplast cell or progeny thereof), and optionally also operably linked to a replication sequence (e.g., a plant origin of replication, a bacterial origin of replication (e.g., for use as a shuttle vector for transferring materials from bacteria to plants), an Agrobacterium Ti plasmid origin of replication, a viral replicon (e.g., for a plant virus), or the like). In a specific embodiment, the invention provides a library of transformed plant protoplasts, or progeny thereof, wherein each transformed protoplast harbors at least one distinct species of heterologous polynucleotide sequence operably linked to an Agrobacterium Ti plasmid in expressible form such that substantially each species of heterologous polynucleotide sequence is transcribed and translated in the host plant cell protoplast or progeny thereof.

In a variation of the embodiment, the heterologous sequences cloned into the Ti plasmid are cDNA sequences obtained from an organism distinct from the phylogenetic species of the plant cell protoplast. In a variation of the embodiment, the heterologous sequences are mutated variants of one or more heterologous parental sequences and/or of one or more sequences present in the genome of the phylogenetic species of the plant cell protoplast; such mutation(s) can be introduced by any suitable method, including but not limited to error-prone PCR, site-directed mutagenesis, oligonucleotide-spiking, or other methods known in the art.

The invention also provides a method for obtaining a desired polynucleotide sequence, comprising selecting, from a population of protoplast library members or their clonal progeny, wherein said protoplast library members each comprise a plant cell protoplast harboring intracellularly one or a subset of a library of

heterologous polynucleotide sequences, a subpopulation of said library members which express a predetermined phenotype. In an aspect, the step of selecting comprises assaying a detectable biochemical phenotype in library members and segregating into a subpopulation those library members which exhibit said detectable biochemical phenotype; typically, the heterologous polynucleotide sequences are recovered from the selected subpopulation. These selected heterologous sequences can be used directly for a variety of uses, can be subjected to one or more subsequent rounds of transformation and selection, and/or can be mutagenized and/or sequence-shuffled and subjected to a subsequent round of transformation and selection, or combinations thereof.

In a broad general aspect, the present invention provides a method for rapid evolution of polynucleotide sequences conferring a desired or predetermined phenotype to at least one plant species, algal species, or cyanobacterium. Typically, the method comprises: (1) transferring a first population of sequence-shuffled polynucleotides comprising a genetic sequence (e.g., a coding sequence, transcriptional or translational regulatory sequence, RNA stability-regulating sequence, etc.) into a plurality of plant cells to produce a first population of transformed plant cells wherein the sequence-shuffled polynucleotides are expressible (either as a coding sequence or as a functional non-coding sequence), either constitutively or conditionally, to confer a phenotype to the transformed plant cell, and optionally to its clonal progeny, (2) selecting, from the first population of transformed plant cells, and/or optionally from clonal progeny thereof, a plurality of genotypes present in said first population of transformed plant cells and expressing the desired phenotype, thereby generating a collection of selected genotypes, (3) producing a second population of sequence-shuffled polynucleotides comprising said genetic sequence obtained (e.g., directly, via in vivo recombination, via amplification, via replication in a shuttle vector, via plant virus transduction, cell fusion, viral superinfection, or after a subsequent manipulation such as mutagenesis, fragmentation, or the like) from the collection of selected genotypes and transferring said second population into a plurality of plant cells forming a second population of transformed plant cells, and optionally clonal progeny thereof, and (4) selecting or identifying from the second population of transformed plant cells at least one genotype present in said second population of transformed plant cells and expressing

the desired phenotype, thereby identifying at least one genotype comprising an evolved shuffled genetic sequence.

Cycles of sequence shuffling, transfer into host cells, and selection typically are repeated iteratively until at least one genetic sequence possesses a satisfactory capacity to produce the desired phenotype; usually from 2 to 1000 cycles of iterative shuffling, transfer into host cells, and selection, with a common range of from 5 to 50 cycles. In one important variation, sequences are recombined recursively prior to selection (either in vitro, or in vivo, e.g., following protoplast fusion), thereby increasing the diversity available for selection.

The number of cycles (with a cycle optionally including multiple rounds of recombination prior to selection) for complete optimization of a genetic sequence depends upon many factors, including the choice of endpoint for optimization, and the skilled artisan is capable of making the determination that a genetic sequence is sufficiently optimized for the intended use and that the recursive evolution can be terminated. In the present invention, at least one cycle of the method comprises transfer into plant cells, such as protoplasts, of shuffled polynucleotides having the genetic sequence(s) to be evolved to confer a desired phenotype, and often at least one cycle comprises selection in plant cell culture, such as by a metabolic assay of cultured plant cells generated from a protoplast transformation, or other selection methodology applicable to plant cell cultures. Once evolved by the method of the present invention, the evolved polynucleotide specie(s) often are transferred into a host organism by any suitable method for transferring the evolved gene into germplasm of a plant species, such as, for example and not limitation, a plant cell protoplast competent for regeneration of an adult organism, which generally may be capable of sexual reproduction and/or asexual propagation by any art-known propagation method. A variation of the method comprises transfer of the evolved genetic sequence into adult plants or plant parts (e.g., a leaf or root) by abrasive transfer (applying the transgene to an abraded surface, with or without an excipient such as LipofectinTM) or biolistics. A variation of the method includes the further step of genetically crossing (e.g., by conventional breeding, protoplast fusion, or recombinant molecular methods) an adult plant harboring an evolved polynucleotide of the invention with a second (or multiple) individual plants, typically of the same species.

An aspect of the invention provides a method for obtaining polynucleotide sequences conferring a desired phenotypic trait to a plant cell, although the method is general and can be used in conjunction with an algal cell or cyanobacterium for certain desired applications. An embodiment of the method comprises transferring into a population of plant cell protoplasts a plurality of library members, wherein library members each comprise a sequence-shuffled polynucleotide obtained by shuffling a plurality of species of a genetic sequence, and selecting from the resultant population of transformed plant cell protoplasts at least one plant cell, or clonal progeny thereof, exhibiting the desired phenotype. Initially, the plurality of species of the genetic sequence that is shuffled are obtained by mutagenesis of one or more starting ("parental") genetic sequence(s), and/or may be obtained from a plurality of parental genetic sequences from nonisogenic individuals of the same or different species (e.g., allogenic - as distinct alleles of a gene locus, or xenogenic - obtained from a plurality of different organismal species and sharing sufficient nucleotide sequence homology for shuffling, or a combination thereof), or alternative sources as is described in commonly-assigned PCT patent applications published as WO98/13487 and WO98/13485 or other related informational publications cited herein.

The invention provides a method for identifying polynucleotide sequences encoding a predetermined phenotype for a plant cell, the method comprises: (1) transforming a plurality of species of sequence-shuffled polynucleotides into protoplasts of plant cells which are clonal progeny of a predetermined non-regenerating plant cell line, and (2) selecting transformed non-regenerable protoplasts or their clonal progeny by segregating individual transformants or pools thereof which express a predetermined phenotype and recovering at least one polynucleotide sequence of a sequence-shuffled polynucleotide. In a variation, the method comprises the further step of culturing the transformed protoplasts on a semisolid medium in growth conditions to form a population of microcalli, wherein substantially each microcallus comprises the clonal progeny of a transformed protoplast; the microcalli or portions thereof are then subjected to selection for the desired phenotype(s). In an aspect, the sequence-shuffled polynucleotides comprise a selectable marker gene and the semisolid medium and/or growth conditions first select for transformants expressing the

selectable marker gene which are capable of growth into microcalli whereas untransformed protoplasts and their progeny are relatively less capable of growth into microcalli. In an aspect, the semisolid protoplast growth medium is M2 and contains an agent which selects for cells expressing a marker gene encoding antibiotic
5 resistance or herbicide resistance. In an alternative embodiment, the transformed protoplasts are propagated as suspensions of callus cells wherein the clonal progeny of individual transformants are propagated in discrete culture vessels; in a specific embodiment the culture vessels are individual wells of a multiwell culture plate.

In a variation, the invention provides a method for isolating novel
10 genetic sequences which confer a predetermined phenotype to a plant cell or plant when expressed therein, the method comprising screening a population of microcalli generated by transforming a population of plant protoplasts with a plurality of library members, wherein library members comprise a sequence-shuffled genetic sequence in expressible form. In an embodiment, the screening comprises performing a
15 biochemical assay on the microcalli or portion thereof. In a specific embodiment, the screening comprises performing a biochemical assay for detecting an enzyme activity; in one variation, the enzyme activity screened for can also be detected in at least one naturally occurring species of the Kingdom Plantae and is encoded by a naturally-occurring plant genome. In a variation, the screening comprises obtaining a cellular
20 sample of each microcallus (or pool) and performing an assay on the cellular sample which utilizes destructive testing of the cellular sample for obtaining readout of the assay. In an alternative embodiment, instead of microcalli, the clonal progeny of the transformed protoplasts are propagated as suspension cultures in liquid protoplast growth medium in discrete culture vessels. In an aspect, the protoplasts used for
25 transformation are obtained from a plant cell line that is predetermined to be non-regenerating, such that adult plants can not be formed under conventional protoplast regeneration conditions.

With regard to the method variations of the invention described herein, the sequence-shuffled polynucleotides can be transformed into protoplasts as naked
30 DNA, as part or all of a genome of a plant virus (encapsidated or as naked nucleic acid), as a lipid-polynucleotide complex, as polynucleotide-coated microprojectiles, or alternative delivery forms known in the art.

The invention also provides a recombinogenic protoplast plant cell suitable for hosting in vivo sequence shuffling, said recombinogenic protoplast plant cell comprising a plant cell which is either stably or transiently transformed with a polynucleotide capable of expressing a recombinase activity which does not naturally occur in the plant species from which the plant cell was derived. For example and not limitation, a recombinogenic plant cell can comprise a cell of a monocot or dicot plant which also has a polynucleotide encoding a bacterial recA recombinase (or a FLP recombinase or cre recombinase for site-specific recombination) in expressible form (e.g., under the transcriptional control of a plant promoter or plant virus promoter functional in said cell, and other variations.) The invention provides a method for performing in vivo sequence shuffling of multiple species of a genetic sequence, the method comprising transforming a population of recombinogenic plant cell protoplasts with a plurality of library members, wherein library members each comprise a polynucleotide species of a genetic sequence, under conditions whereby greater than about 2 percent, preferably more than about 5 percent to about 10 percent or more, of the transformed recombinogenic plant cell protoplasts are co-transformed with multiple species of library members and expressed recombinase activity facilitates homologous or site-specific recombination between library members within the plant cell protoplast or its clonal progeny, and culturing the resultant co-transformed protoplasts and progeny. In a variation, the encoded recombinase is inducible, such as by being operably linked to an inducible promoter which can be induced in a plant cell by application of induction conditions. In a variation, the method comprises the further step of selecting from the resultant population of co-transformed plant cell protoplasts at least one plant cell, or clonal progeny thereof, exhibiting the desired phenotype. Optionally, the in vivo shuffled library members can be recovered for subsequent transformation into plant cell protoplasts (recombinogenic or non-recombinogenic), either prior to or subsequent to a phenotype selection step.

The invention provides a plant cell protoplast and clonal progeny thereof containing a sequence-shuffled polynucleotide which is not encoded by the naturally occurring genome of the plant cell protoplast. The invention also provides a collection of plant cell protoplasts transformed with a library of sequence-shuffled polynucleotides in expressible form. The invention further provides a plant cell

protoplast co-transformed with at least two species of library members wherein library members comprise sequence-shuffled polynucleotides encoding a genetic sequence.

5 The invention also provides a regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled portion, typically in expressible form. The invention provides a method variation wherein at least one round of phenotype selection is performed on regenerated plants derived from protoplasts transformed with sequence-shuffled library members.

10 The present invention provides a method for generating polynucleotide sequences encoding at least one novel or modified phenotype which can be selected on the basis of expression of a genetic sequence in a plant protoplast, plant cell culture, or organism regenerated from a plant protoplast. Although not intended to be an exhaustive list, the following illustrative examples of such phenotypes include: a biosynthetic enzyme, a multi-enzyme biosynthetic pathway,
15 enzymatic activity, resistance to insect infestation, resistance to a plant pathogen, morphological characteristic, foodstuff content, flavor component, altered fruit ripening, vegetative growth, senescence, carbon-fixation rate, nitrogen fixation, interaction with Rhizobium and/or other microbes, photosynthetic efficiency, herbicide resistance, pesticide resistance, flowering, photoperiodism, shelf-life,
20 growth rate, growth habit, starch content, protein content, frost resistance, pigment content, nutrient content, genes encoding functions that effect transformation efficiency and efficient somatic regeneration, and the like. The phenotype modification can result from introduction of an optimized gene, gene fragment, or regulatory sequence derived from a genome of a plant (e.g., from a genome of an organism in the Kingdom Plantae), a plant virus genome, a microorganism genome
25 (including episomal vectors thereof), an animal genome, an animal virus genome, or a combination thereof. The optimized gene, gene fragment, or regulatory sequence is obtained by recursive sequence shuffling which is described further herein and in documents incorporated herein by reference. The recursive sequence shuffling is typically employed to obtain and/or optimize function of the gene, gene product, gene
30 segment, and/or regulatory sequence in a plant host, in a prokaryotic host that is suitable for agricultural use (e.g., *Agrobacterium tumefaciens*, and *ice*⁽⁻⁾ leaf commensal bacteria, etc.), or in a plant virus. An important aspect of the present

invention is that the method employs at least one step wherein sequence-shuffled polynucleotides are introduced into plant cell protoplasts to produce a library of transformed protoplasts which can be selected for the presence of a desired predetermined phenotype, either directly or by performing selection on clonal
5 progeny of the transformed protoplast.

The invention provides a kit for obtaining a polynucleotide encoding a predetermined phenotype, the kit comprising a plant cell line suitable for forming transformable protoplasts and a collection sequence-shuffled polynucleotides formed by in vitro sequence shuffling. The kit often further comprises a transformation
10 enhancing agent (e.g., lipofection agent, PEG, etc.) and/or a transformation device (e.g., a biolistics gene gun) and/or a plant viral vector which can infect plant cells or protoplasts thereof. The kit also optionally comprises buffers, containers, packaging materials, instructions for practicing the methods herein, or the like.

Although the methods of the invention are believed to be suitable for
15 use with substantially any plant type, including gymnosperms, angiosperms (including dicots and monocots), ferns, and algae, it is described with particular reference to higher plants for illustrative purposes.

The disclosed method for altering a agricultural organism phenotype by iterative gene shuffling and phenotype selection is a pioneering method which
20 enables a broad range of novel and advantageous agricultural compositions, methods, kits, uses, plant cultivars, and apparatus which will be apparent to those skilled in the art in view of the present disclosure.

Other features and advantages of the invention will be apparent from the following description of the drawings, preferred embodiments of the invention,
25 the examples, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic portrayal of a generic plasmid for transduction/transformation of cloned heterologous polynucleotide sequences into cells.

DETAILED DESCRIPTION

GENERAL OVERVIEW

The present invention provides methods, reagents, genetically modified plants, plant cells and protoplasts thereof, microbes (e.g., *Agrobacterium*),
5 polynucleotides, shuffled nucleic acids, other protoplasts (such as fungal protoplasts),
plant cell and plant libraries, fungal cells and fungal organism libraries and
compositions relating to the forced evolution of genetic sequences that confer
selectable phenotypes to agricultural organisms, or portions thereof, having a desired
phenotypic alteration generated by polynucleotide sequence shuffling of a plurality of
10 polynucleotide sequences, typically having regions of substantial sequence identity to
facilitate shuffling recombinations. For example, the invention provides methods and
related compositions for introducing libraries of shuffled nucleic acids into plant
protoplasts, and selecting the protoplasts (or corresponding regenerated plant cells or
plants) for a desired trait or property. Nucleic acids from the plants or protoplasts can
15 be isolated to produce secondary libraries which can be transduced into cells or
protoplasts, which are again selected for a desired trait or property. This process can
be repeated one, two, three, four or more times until a desired trait or property is
obtained.

Similarly, plants, cells or protoplasts which are selected can be
20 transduced with one or more additional library of nucleic acids, which recombine in
the plants, cells or protoplasts, and which are selected for a desired trait or property.
This process can also be repeated one or several times and multiple cycles of
recombination can be performed prior to selection (or between rounds of selection) to
increase the diversity available during screening stages. Libraries of materials can be
25 shuffled nucleic acids produced by any available shuffling methodology, or can be
focused or random libraries of nucleic acids. In either case, the nucleic acids of the
libraries can remain unrecombined in cells or protoplasts into which the nucleic acids
are transduced, or the nucleic acids can recombine with nucleic acids previously
present in the cells, plants or protoplasts (e.g., genomic or episomal DNAs). To aid in
30 recombination, plant cells, plants or protoplasts can be transduced with genes which
encode recombinogenic proteins (such as *recA*), or libraries of materials can be coated
with the recombinogenic proteins themselves (e.g., the *recA* protein). Commonly,
transduction of recombinogenic factors (nucleic acids, proteins or other materials) is

performed at the same time as transduction with the library of interest. Nucleic acids can be present in transducing vectors such as Agrobacterium vectors which facilitate recombination of sequences of interest into host DNAs (e.g., genomic DNAs).

Definitions

5 Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of
10 the present invention, the following terms are defined below.

The term "destructive testing" is defined herein as a procedure to determine a biochemical, biophysical, genetic, or other property or parameter of a plant cell or protoplast, which procedure results in the assayed cells thereby becoming non-replicable and/or non-viable. For example and not limitation, destructive testing
15 can include assays which use cell lysis (irreparable damage to the cell membrane and/or cell wall), exposure to genotoxic or toxic chemicals, ionizing or ultraviolet irradiation at flux levels sufficient to lethally damage the irradiated cells, and the like.

The term "derivative" refers to a component (e.g., a library of molecules) made using a specified parental (e.g., an original library of molecules)
20 component.

The term "reassembly" is used when recombination occurs between identical polynucleotide sequences.

By contrast, the term "shuffling" is used herein to indicate recombination between substantially homologous but non-identical polynucleotide
25 sequences. In some embodiments, DNA shuffling may involve crossover via nonhomologous recombination, such as via cre/lox and/or flp/frt systems or via oligonucleotide or in silico shuffling, or the like, such that recombination need not require substantially similar polynucleotide sequences. Homologous and non-homologous recombination formats can be used, and, in some embodiments, can
30 generate molecular chimeras and/or molecular hybrids of substantially dissimilar sequences. Viral recombination systems, such as template-switching and the like can also be used to generate molecular chimeras and recombined genes, or portions thereof. A general description of shuffling is provided in commonly-assigned

WO98/13487 and WO98/13485, both of which are incorporated herein in their entirety by reference; in case of any conflicting description of definition between any of the incorporated documents and the text of this specification, the present specification provides the principal basis for guidance and disclosure of the present invention.

The term "related polynucleotides" means that regions or areas of the polynucleotides at issue are identical and regions or areas of the polynucleotides are heterologous.

The term "chimeric polynucleotide" means that the polynucleotide comprises regions which are wild-type and regions which are mutated, or that the polynucleotide has nucleic acid subsequences derived from more than one source, depending on the context herein. It can also mean that the polynucleotide comprises wild-type regions from one polynucleotide and wild-type regions from another related polynucleotide.

The term "cleaving" means digesting the polynucleotide with enzymes or breaking the polynucleotide (e.g., by chemical or physical means), or generating partial length copies of a parent sequence(s) via partial PCR extension, PCR stuttering, differential fragment amplification, or other means of producing partial length copies of one or more parental sequences.

The term "population" as used herein means a collection of components such as polynucleotides, nucleic acid fragments or proteins. A "mixed population" means a collection of components which belong to the same family of nucleic acids or proteins (i.e. are related) but which differ in their sequence (i.e. are not identical) and hence in their biological activity.

The term "mutations" means changes in the sequence of a parent nucleic acid sequence (e.g., a gene or a microbial genome, transferable element, or episome) or changes in the sequence of a parent polypeptide. Such mutations may be point mutations such as transitions or transversions. The mutations may be deletions, insertions or duplications.

The term "recursive sequence recombination" as used herein refers to a method whereby a population of polynucleotide sequences are recombined with each other by any suitable recombination means (e.g., sexual PCR, homologous recombination, site-specific recombination, etc.) to generate a library of sequence-

recombined species which is then screened or subjected to selection to obtain those sequence-recombined species having a desired property; the selected species are then subjected to at least one additional cycle of recombination with themselves and/or with other polynucleotide species and at subsequent selection or screening for the
5 desired property.

The term "amplification" means that the number of copies of a nucleic acid fragment is increased.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or
10 polynucleotide sequence that is present in an organism that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. As used herein, laboratory strains and established cultivars of plants which may have been selectively bred according to classical genetics are considered naturally-occurring. As used herein, naturally-occurring
15 polynucleotide and polypeptide sequences are those sequences, including natural variants thereof, which can be found in a source in nature, or which are sufficiently similar to known natural sequences that a skilled artisan would recognize that the sequence could have arisen by natural mutation and recombination processes.

As used herein "predetermined" means that the cell type, non-human
20 animal, or virus may be selected at the discretion of the practitioner on the basis of a known phenotype.

As used herein, "linked" means in polynucleotide linkage (i.e., phosphodiester linkage). "Unlinked" means not linked to another polynucleotide sequence; hence, two sequences are unlinked if each sequence has a free 5' terminus
25 and a free 3' terminus.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding
30 sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where appropriate to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases

and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous. A structural gene (e.g., a *ESPSP* gene) which is operably linked to a polynucleotide sequence corresponding to a transcriptional regulatory sequence of an endogenous gene is generally expressed in substantially the same temporal and cell type-specific pattern as is the naturally-occurring gene.

As used herein, the terms "expression cassette" refers to a polynucleotide comprising a promoter sequence and, optionally, an enhancer and/or silencer element(s), operably linked to a structural sequence, such as a cDNA sequence or genomic DNA sequence. In some embodiments, an expression cassette may also include polyadenylation site sequences to ensure polyadenylation of transcripts. When an expression cassette is transferred into a suitable host cell, the structural sequence is transcribed from the expression cassette promoter, and a translatable message is generated, either directly or following appropriate RNA splicing. Typically, an expression cassette comprises: (1) a promoter, such as a CaMV 35S promoter, a NOS promoter or a *rbcS* promoter, or other suitable promoter known in the art, (2) a cloned polynucleotide sequence, such as a cDNA or genomic fragment ligated to the promoter in sense orientation so that transcription from the promoter will produce a RNA that encodes a functional protein, and (3) a polyadenylation sequence. For example and not limitation, an expression cassette of the invention may comprise the cDNA expression cloning vectors, pCD and λ NMT (Okayama H and Berg P (1983) Mol. Cell. Biol. 3: 280; Okayama H and Berg P (1985) Mol. Cell. Biol. 5: 1136, incorporated herein by reference).

The term "transcriptional modulation" is used herein to refer to the capacity to either enhance transcription or inhibit transcription of a structural sequence linked in cis; such enhancement or inhibition may be contingent on the occurrence of a specific event, such as stimulation with an inducer and/or may only be manifest in certain cell types. The altered ability to modulate transcriptional enhancement or inhibition may affect the inducible transcription of a gene or may effect the basal level transcription of a gene, or both. Numerous other specific examples of transcription regulatory elements, such as specific enhancers and silencers, are known to those of skill in the art and may be selected for use in the methods and polynucleotide constructs of the invention on the basis of the practitioner's desired application. Literature sources and published patent documents,

as well as GenBank™ and other sequence information data sources can be consulted by those of skill in the art in selecting suitable transcription regulatory elements for use in the invention. Where appropriate, a transcription regulatory element may be constructed by synthesis (and ligation, if necessary) of oligonucleotides made on the basis of available sequence information (e.g., GenBank sequences).

As used herein, the term "transcriptional unit" or "transcriptional complex" refers to a polynucleotide sequence that comprises a structural gene (exons), a cis-acting linked promoter and other cis-acting sequences necessary for efficient transcription of the structural sequences, distal regulatory elements necessary for appropriate tissue-specific and developmental transcription of the structural sequences, and additional cis sequences important for efficient transcription and translation (e.g., polyadenylation site, mRNA stability controlling sequences).

As used herein, the term "transcription regulatory region" refers to a DNA sequence comprising a functional promoter and any associated transcription elements (e.g., enhancer, CCAAT box, TATA box, LRE, ethanol-inducible element, etc.) that are essential for transcription of a polynucleotide sequence that is operably linked to the transcription regulatory region.

As used herein, the term "xenogeneic" is defined in relation to a recipient genome, host cell, or organism and means that an amino acid sequence or polynucleotide sequence is not encoded by or present in, respectively, the naturally-occurring genome of the recipient genome, host cell, or organism. Xenogenic DNA sequences are foreign DNA sequences. Further, a nucleic acid sequence that has been substantially mutated (e.g., by site directed mutagenesis) is xenogeneic with respect to the genome from which the sequence was originally derived, if the mutated sequence does not naturally occur in the genome.

As used herein, the term "minigene" or "minilocus" refers to a heterologous gene construct wherein one or more nonessential segments of a gene are deleted with respect to the naturally-occurring gene. Typically, deleted segments are intronic sequences of at least about 100 basepairs to several kilobases, and may span up to several tens of kilobases or more. Isolation and manipulation of large (i.e., greater than about 50 kilobases) targeting constructs is frequently difficult and may reduce the efficiency of transferring the targeting construct into a host cell. Thus, it is frequently desirable to reduce the size of a targeting construct by deleting one or more

nonessential portions of the gene. Typically, intronic sequences that do not encompass essential regulatory elements may be deleted. Frequently, if convenient restriction sites bound a nonessential intronic sequence of a cloned gene sequence, a deletion of the intronic sequence may be produced by: (1) digesting the cloned DNA
5 with the appropriate restriction enzymes, (2) separating the restriction fragments (e.g., by electrophoresis), (3) isolating the restriction fragments encompassing the essential exons and regulatory elements, and (4) ligating the isolated restriction fragments to form a minigene wherein the exons are in the same linear order as is present in the germline copy of the naturally-occurring gene. Alternate methods for producing a
10 minigene will be apparent to those of skill in the art (e.g., ligation of partial genomic clones which encompass essential exons but which lack portions of intronic sequence). Most typically, the gene segments comprising a minigene will be arranged in the same linear order as is present in the germline gene, however, this will not always be the case. Some desired regulatory elements (e.g., enhancers, silencers) may
15 be relatively position-insensitive, so that the regulatory element will function correctly even if positioned differently in a minigene than in the corresponding germline gene. For example, an enhancer may be located at a different distance from a promoter, in a different orientation, and/or in a different linear order. For example, an enhancer that is located 3' to a promoter in germline configuration might be located
20 5' to the promoter in a minigene. Similarly, some genes may have exons which are alternatively spliced at the RNA level, and thus a minigene may have fewer exons and/or exons in a different linear order than the corresponding germline gene and still encode a functional gene product. A cDNA encoding a gene product may also be used to construct a minigene. However, since it is generally desirable that the
25 heterologous minigene be expressed similarly to the cognate naturally-occurring nonhuman gene, transcription of a cDNA minigene typically is driven by a linked gene promoter and enhancer from the naturally-occurring gene.

The term "corresponds to" is used herein to mean that a polynucleotide sequence is identical or complementary to all or a portion of a reference
30 polynucleotide sequence, or that a polypeptide sequence is identical to at least a substantial portion of a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For

illustration, the nucleotide sequence "5'-TATAC" corresponds to a reference sequence "5'-TATAC" and is complementary to a reference sequence "5'-GTATA".

"Physiological conditions" as used herein refers to temperature, pH, ionic strength, viscosity, and like biochemical parameters that are compatible with a viable plant organism or agricultural microorganism (e.g., *Rhizobium*, *Agrobacterium*, etc.), and/or that typically exist intracellularly in a viable cultured plant cell, particularly conditions existing in the nucleus of said cell. In general, in vitro physiological conditions can comprise 50-200 mM NaCl or KCl, pH 6.5-8.5, 20-45°C and 0.001-10 mM divalent cation (e.g., Mg^{++} , Ca^{++}); preferably about 150 mM NaCl or KCl, pH 7.2-7.6, 5 mM divalent cation, and often include 0.01-1.0 percent nonspecific protein (e.g., BSA). A non-ionic detergent (Tween, NP-40, Triton X-100) can often be present, usually at about 0.001 to 2%, typically 0.05-0.2% (v/v). Particular aqueous conditions may be selected by the practitioner according to conventional methods. For general guidance, the following buffered aqueous conditions may be applicable: 10-250 mM NaCl, 5-50 mM Tris HCl, pH 5-8, with optional addition of divalent cation(s), metal chelators, nonionic detergents, membrane fractions, antifoam agents, and/or scintillants.

As used herein, the terms "label" or "labeled" refer to incorporation of a detectable marker, e.g., a radiolabeled amino acid or a recoverable label (e.g. biotinyl moieties that can be recovered by avidin or streptavidin). Recoverable labels can include covalently linked polynucleobase sequences that can be recovered by hybridization to a complementary sequence polynucleotide. Various methods of labeling polypeptides, PNAs, and polynucleotides are known in the art and may be used. Examples of labels include, but are not limited to, the following: radioisotopes (e.g., 3H , ^{14}C , ^{35}S , ^{125}I , ^{131}I), fluorescent or phosphorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), enzymatic labels (e.g., horseradish peroxidase, β -galactosidase, luciferase, alkaline phosphatase), biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths, e.g., to reduce potential steric hindrance.

As used herein, the term "statistically significant" means a result (i.e., an assay readout) that generally is at least two standard deviations above or below the

mean of at least three separate determinations of a control assay readout and/or that is statistically significant as determined by Student's t-test or other art-accepted measure of statistical significance.

5 The term "transcriptional modulation" is used herein to refer to the capacity to either enhance transcription or inhibit transcription of a structural sequence linked in cis; such enhancement or inhibition may be contingent on the occurrence of a specific event, such as stimulation with an inducer and/or may only be manifest in certain cell types.

10 The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal cells or tissues. Agents are evaluated for potential activity as antiviral agents by inclusion in screening assays described hereinbelow.

15 As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual macromolecular species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 to 90 percent of all
20 macromolecular species present in the composition. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered
25 macromolecular species.

As used herein, the term "optimized" is used to mean substantially improved in a desired structure or function relative to an initial starting condition, not necessarily the optimal structure or function which could be obtained if all possible combinatorial variants could be made and evaluated, a condition which is typically
30 impractical due to the number of possible combinations and permutations in polynucleotide sequences of significant length (e.g., a complete plant gene or genome).

As used herein, "phenotype" means an observable or otherwise detectable manifestation of a heritable trait or encoded function. For example and not limitation, a phenotype can comprise an enzyme activity, a metabolic pathway that produces a detectable product or depletes a detectable substrate. A phenotype can
5 comprise a detectable change in the rate of uptake or production of a metabolite, insect resistance, herbicide resistance, and other detectable manifestations of gene expression.

TRANSDUCTION, CLONING AND MOLECULAR BIOLOGY

The procedures herein involve, e.g., making libraries of nucleic acids
10 and transducing protoplasts with the libraries. More generally, the nomenclature used hereafter and the laboratory procedures in agriculture, cell culture (especially plant cell culture), molecular genetics, virology (e.g., of plant viruses and virus-based vectors), and nucleic acid chemistry and hybridization described below are those well known and commonly employed in the art. Standard techniques are used for
15 recombinant nucleic acid methods, polynucleotide synthesis, and microbial culture and transformation (e.g., biolistics, Agrobacterium (Ti plasmid), electroporation, lipofection).

Generally, enzymatic reactions and purification steps are performed according to the manufacturer's specifications. The techniques and procedures are
20 generally performed according to conventional methods in the art and various general references (see, generally, Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al. Molecular Cloning: A Laboratory Manual, 2d ed., Vol. 1-3 (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., which is
25 incorporated herein by reference and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel")) which are provided throughout this document. The general procedures therein are believed to be well known in the art and are provided for the convenience
30 of the reader.

In addition to Berger Ausubel and Sambrook, useful general references for plant cell cloning, culture and regeneration include Payne et al. (1992) Plant Cell

and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); and Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture: Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) (Gamborg). Cell culture media are described in Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL (Atlas).
 5 Additional information is found in commercial literature such as the Life Science Research Cell Culture catalogue (1998) from Sigma- Aldrich, Inc (St Louis, MO) (Sigma-LSRCCC) and, e.g., the Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-PCCS).

10 Oligonucleotides can be synthesized on an Applied Bio Systems oligonucleotide synthesizer according to specifications provided by the manufacturer, or by a variety of other known techniques, or can be ordered from any of a variety of sources, including, e.g., Operon Technologies (Alameda, CA). *See also*, Beaucage and Caruthers (1981), Tetrahedron Letts., 22(20):1859-1862 and
 15 Needham-VanDevanter *et al.* (1984) Nucleic Acids Res., 12:6159-6168.

Methods for PCR amplification are described in the art (PCR Technology: Principles and Applications for DNA Amplification ed. HA Erlich, Freeman Press, New York, NY (1992); PCR Protocols: A Guide to Methods and Applications, eds. Innis, Gelfand, Snisky, and White, Academic Press, San Diego,
 20 CA (1990); Mattila *et al.* (1991) Nucleic Acids Res. 19: 4967; Eckert, K.A. and Kunkel, T.A. (1991) PCR Methods and Applications 1: 17; PCR, eds. McPherson, Quirk, and Taylor, IRL Press, Oxford; and U.S. Patent 4,683,202, which are incorporated herein by reference). Leaf PCR is suitable for genotype analysis of transgene plants.

25 All sequences referred to herein by GenBank database file designation or a commonly used reference name which is indexed in GenBank or otherwise published are incorporated herein by reference and are publicly available.

FORMATS FOR SEQUENCE RECOMBINATION

30 The methods of the invention entail performing recombination ("shuffling") and screening or selection to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes (Stemmer (1995) *Bio/Technology* 13:549-553). This recombination can occur before or after

introduction of nucleic acids into, e.g., plant protoplasts. Reiterative cycles of recombination and screening/selection can be performed to further evolve the nucleic acids of interest. Such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide and genetic engineering. Shuffling allows the recombination of large numbers of mutations in a minimum number of selection cycles, in contrast to natural pairwise recombination events (e.g., as occur during sexual replication). Thus, the sequence recombination techniques described herein provide particular advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence recombination, provides opportunities for modification of the technique.

A number of publications by the inventors and their co-workers describe DNA shuffling, which can be used in the context of the present invention, e.g., to produce libraries of shuffled materials which are transduced into plant protoplasts or cells. For example, Stemmer et al. (1994) "Rapid Evolution of a Protein" Nature 370:389-391; Stemmer (1994) "DNA Shuffling by Random Fragmentation and Reassembly: in vitro Recombination for Molecular Evolution," Proc. Natl. Acad. USA 91:10747-10751; Stemmer U.S. Patent No. 5,603,793 METHODS FOR IN VITRO RECOMBINATION; Stemmer et al. U.S. Pat. No. 5,830,721 DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY and Stemmer et al. U.S. Pat. No. 5,811,238 METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION describe e.g., in vitro protein shuffling methods, e.g., by repeated cycles of mutagenesis, shuffling and selection as well as a variety of methods of generating libraries of displayed peptides and antibodies and a variety of DNA reassembly techniques following DNA fragmentation, and their application to mutagenesis *in vitro* and *in vivo*.

Applications of DNA shuffling technology have also been developed by the inventors and their co-workers, and these methods can be applied to the present invention for library generation and/or screening methodologies. In addition to the publications noted above, Minshull et al., U.S. Pat. No. 5,837,458 METHODS AND

COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING provides for the evolution of new metabolic pathways and the enhancement of bio-processing through recursive shuffling techniques. Cramer et al. (1996), "Construction And Evolution Of Antibody-Phage Libraries By DNA Shuffling" Nature Medicine 2(1):100-103 describe antibody shuffling for antibody phage libraries. Additional details regarding DNA Shuffling can also be found in WO95/22625, WO97/ 20078, WO96/33207, WO97/33957, WO98/27230, WO97/35966, WO98/ 31837, WO98/13487, WO98/13485 and WO98/42832.

A number of the publications of the inventors and their co-workers, as well as other investigators in the art also describe techniques which facilitate DNA shuffling, e.g., by providing for reassembly of genes from small fragments of genes, or even oligonucleotides encoding gene fragments. For example, in addition to the publications noted above, Stemmer et al. (1998) U.S. Pat. No. 5,834,252 END COMPLEMENTARY POLYMERASE REACTION describe processes for amplifying and detecting a target sequence (e.g., in a mixture of nucleic acids), as well as for assembling large polynucleotides from fragments.

CREATION OF RECOMBINANT LIBRARIES

The invention involves creating recombinant libraries of polynucleotides that are then screened to identify those library members that exhibit a desired property. The recombinant libraries can be created using any of the various methods herein, as well as many others which would be apparent to one of skill.

Methods for obtaining recombinant polynucleotides and/or for obtaining diversity in nucleic acids, e.g., as in molecular libraries of such polynucleotides, e.g., used as the substrates for DNA shuffling as described herein include, for example, homologous recombination (e.g., PCT/US98/05223; Publ. No. WO98/42727 and the other references noted herein); oligonucleotide-directed mutagenesis (for review see, Smith, *Ann. Rev. Genet.* 19: 423-462 (1985); Botstein and Shortle, *Science* 229: 1193-1201 (1985); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)). Included among these methods are oligonucleotide-directed mutagenesis (Zoller and Smith, *Nucl. Acids Res.* 10: 6487-6500 (1982), *Methods in Enzymol.* 100: 468-500

(1983), and *Methods in Enzymol.* 154: 329-350 (1987)) phosphothioate-modified DNA mutagenesis (Taylor *et al.*, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor *et al.*, *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye and Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Sayers *et al.*, *Nucl. Acids Res.* 16: 791-802 (1988); Sayers *et al.*, *Nucl. Acids Res.* 16: 803-814 (1988)), mutagenesis using uracil-containing templates (Kunkel, *Proc. Nat'l. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel *et al.*, *Methods in Enzymol.* 154: 367-382)); mutagenesis using gapped duplex DNA (Kramer *et al.*, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer and Fritz, *Methods in Enzymol.* 154: 350-367 (1987); Kramer *et al.*, *Nucl. Acids Res.* 16: 7207 (1988)); and Fritz *et al.*, *Nucl. Acids Res.* 16: 6987-6999 (1988)). Additional suitable methods include point mismatch repair (Kramer *et al.*, *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Carter *et al.*, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, *Methods in Enzymol.* 154: 382-403 (1987)), deletion mutagenesis (Eghtedarzadeh and Henikoff, *Nucl. Acids Res.* 14: 5115 (1986)), restriction-selection and restriction-purification (Wells *et al.*, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986)), mutagenesis by total gene synthesis (Nambiar *et al.*, *Science* 223: 1299-1301 (1984); Sakamar and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells *et al.*, *Gene* 34: 315-323 (1985); and Grundström *et al.*, *Nucl. Acids Res.* 13: 3305-3316 (1985). Kits for mutagenesis are commercially available (e.g., Bio-Rad, Amersham International, Anglian Biotechnology).

In a presently preferred embodiment, the recombinant libraries are prepared using DNA shuffling. The shuffling and screening or selection can be used to "evolve" individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes (Stemmer (1995) *Bio/Technology* 13:549-553).

Reiterative cycles of recombination and screening/selection can be performed to further evolve the nucleic acids of interest. These cycles can occur before or after transduction of libraries into protoplasts, and multiple cycles of recombination can be performed prior to cycles of selection (conversely, especially where a population is highly diverse and selective forces are weak, multiple cycles of selection can be performed between cycles of recombination). In general, such techniques do not require the extensive analysis and computation required by conventional methods for polypeptide engineering. Shuffling allows the recombination of large numbers of mutations in a minimum number of selection

cycles, in contrast to traditional, pairwise recombination events. Thus, the sequence recombination techniques described herein provide particular advantages in that they provide recombination between mutations in any or all of these, thereby providing a very fast way of exploring the manner in which different combinations of mutations can affect a desired result. In some instances, however, structural and/or functional information is available which, although not required for sequence recombination, provides opportunities for modification of the technique.

As noted above, exemplary formats and examples for sequence recombination, sometimes referred to as DNA shuffling, evolution, or molecular breeding, have been described by the present inventors and co-workers in co-pending applications and can be applied to the present invention for generating libraries which are transduced into plant or fungal protoplasts for screening (or for screening in regenerated cells, plants or fungi). For example, U.S. Patent Application Serial No. 08/198,431, filed February 17, 1994, Serial No. PCT/US95/02126, filed, February 17, 1995, Serial No. 08/425,684, filed April 18, 1995, Serial No. 08/537,874, filed October 30, 1995, Serial No. 08/564,955, filed November 30, 1995, Serial No. 08/621,859, filed March 25, 1996, Serial No. 08/621,430, filed March 25, 1996, Serial No. PCT/US96/05480, filed April 18, 1996, Serial No. 08/650,400, filed May 20, 1996, Serial No. 08/675,502, filed July 3, 1996, Serial No. 08/721, 824, filed September 27, 1996, Serial No. PCT/US97/17300, filed September 26, 1997, and Serial No. PCT/US97/24239, filed December 17, 1997; Stemmer, *Science* 270:1510 (1995); Stemmer *et al.*, *Gene* 164:49-53 (1995); Stemmer, *Bio/Technology* 13:549-553 (1995); Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91:10747-10751 (1994); Stemmer, *Nature* 370:389-391 (1994); Cramer *et al.*, *Nature Medicine* 2(1):1-3 (1996); Cramer *et al.*, *Nature Biotechnology* 14:315-319 (1996) each teach nucleic acid recombination and shuffling methods applicable to the present invention which can be used for library generation.

Further in this regard, the following co-pending patent applications and publications of the present inventors and co-workers are incorporated herein by reference for all purposes: U.S.S.N. 08/198,431, filed 17 February 1994, PCT/US95/02126 filed 17 February 1995, WO97/20078, U.S. Patent 5,605,793, U.S. Patent 5,358,665, U.S. Patent 5,270,170, U.S.S.N. 08/425,684 filed 18 April 1995, U.S.S.N. 08/537,874 filed 30 October 1995, U.S.S.N. 08/564,955 filed 30 November

1995, U.S.S.N. 08/621,859 filed 25 March 1996, PCT/US96/05480 filed 18 April 1996, U.S.S.N. 08/650,400 filed 20 May 1996, U.S.S.N. 08/675,502 filed 3 July 1996, U.S.S.N. 08/721,824 filed 27 September 1996, U.S.S.N. 08/722,660 filed 27 September 1996, and U.S.S.N. 08/769,062 filed 18 December 1996; WO98/13485 and WO98/13487; and Stemmer (1995) Science 270: 1510; Stemmer et al. (1995) Gene 164: 49-53; Stemmer (1995) Bio/Technology 13: 549-553; Stemmer (1994) PNAS 91: 10747-10751; Stemmer (1994) Nature 370: 389-391; Cramer et al. (1996) Nature Medicine 2: 1-3; Cramer et al. (1996) Nature Biotechnology 14: 315-319.

Additional Application of Shuffling Technologies to the Invention--
An Overview

The invention relates in part to a generally applicable method for generating novel or improved agricultural organisms (e.g., plants or fungi) or genetic sequences relating thereto comprising genotypes and phenotypes which do not naturally occur or would be anticipated to occur at a substantial frequency in nature.

A broad aspect of the method employs recursive nucleotide sequence recombination, termed "sequence shuffling", which enables the rapid generation of a collection of broadly diverse phenotypes that can be selectively bred for a broader range of novel phenotypes or more extreme phenotypes than would otherwise occur by natural evolution in the same time period. A basic variation of the method is a recursive process comprising: (1) sequence shuffling of a plurality of species of a genetic sequence, which species may differ by as little as a single nucleotide difference or may be substantially different, yet retain sufficient regions of sequence similarity or site-specific recombination junction sites to support shuffling recombination (this step is optionally reiterated before performing step 2, or can be repeatedly performed on material selected in step 2); (2) selection of the resultant shuffled genetic sequence to isolate or enrich a plurality of shuffled genetic sequences having a desired phenotype(s) (this step is also optionally reiterated); and (3) repeating steps (1) and (2) on the plurality of shuffled genetic sequences having the desired phenotype(s) until one or more variant genetic sequences encoding a sufficiently optimized desired phenotype is obtained. In alternative formats, oligonucleotide mediated shuffling, or "in silico" formats are used to generate shuffled libraries.

In these general ways, the methods herein facilitate the "forced evolution" of a novel or improved genetic sequence to encode a desired phenotype

which natural selection and evolution has heretofore not generated in the reference agricultural organism. Shuffling and selection steps can be performed prior to introduction of materials into protoplasts, or subsequent to introduction of materials into protoplasts, or both.

5 Typically, a plurality of genetic sequences of the same gene locus from the same taxonomic classification of organism are shuffled and selected by the present method. A common use of the method is to shuffle mutant variants of a genetic sequence of a plant or fungal genome or a genetic sequence of a microorganism which may function in a plant or fungus, to obtain a variant of the genetic sequence that
10 possesses a novel desired phenotype or an improved desired phenotype. However, the method can be used with a plurality of alleles, homologs, or cognate genes of a genetic locus, or even with a plurality of genetic sequences from related organisms, and in some instances with unrelated genetic sequences or portions thereof which have recombinogenic portions (either naturally or generated via genetic engineering or via
15 in silico or oligonucleotide-mediated recombination methods). Furthermore, the method can be used to evolve a heterologous sequence (e.g., a non-naturally occurring mutant gene) to optimize its phenotypic expression (e.g., function) in a particular genomic background, and/or in a particular host cell or expression system (e.g., an expression cassette or expression replicon).

20 A basic element of the methods herein, termed sequence shuffling (or simply "shuffling"), in broad application, consists of a method for generating a selected polynucleotide sequence or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess or encode a desired phenotypic characteristic
25 (e.g., encode a polypeptide, promote transcription of linked polynucleotides, modify transformation efficiency, bind a protein, and the like) which can be selected for. One method of identifying polypeptides that possess a desired structure or functional property, such as encoding a desired enzymatic function(s) (e.g., an enhanced Rubisco, a herbicide catabolizing enzyme, an optimized plant biosynthetic pathway),
30 involves the screening of a large library of polynucleotides for individual library members which possess or encode the desired structure or functional property conferred by the polynucleotide sequence.

In a general aspect, the invention provides a method, termed "sequence shuffling", for generating libraries of recombinant polynucleotides having a desired characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related-sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. In the method, at least two species of the related-sequence polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence polynucleotide with at least one adjacent portion of at least one second species of a related-sequence polynucleotide. Recombination systems suitable for generating sequence-recombined polynucleotides can be either: (1) in vitro systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) in vivo systems for homologous recombination or site-specific recombination as described herein. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The selected sequence-recombined polynucleotides, which are typically related-sequence polynucleotides, can then be subjected to at least one recursive cycle wherein at least one selected sequence-recombined polynucleotide is combined with at least one distinct species of related-sequence polynucleotide (which may itself be a selected sequence-recombined polynucleotide) in a recombination system suitable for generating sequence-recombined polynucleotides, such that additional generations of sequence-recombined polynucleotide sequences are generated from the selected sequence-recombined polynucleotides obtained by the selection or screening method employed. In this manner, recursive sequence recombination generates library members which are sequence-recombined polynucleotides possessing desired characteristics. Such characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene

or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property.

Nucleic acid sequence shuffling is a method for recursive *in vitro* or *in vivo* homologous or nonhomologous recombination of pools of nucleic acid fragments or polynucleotides (e.g., genes from agricultural organisms or portions thereof).
5 Mixtures of related nucleic acid sequences or polynucleotides are randomly or pseudorandomly fragmented, and reassembled to yield a library or mixed population of recombinant nucleic acid molecules or polynucleotides.

The present invention is directed to a method for generating a selected polynucleotide sequence (e.g., a plant gene or microbe gene, or combinations thereof)
10 or population of selected polynucleotide sequences, typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequence(s) possess a desired phenotypic characteristic (e.g., encode a polypeptide, promote transcription of linked polynucleotides, bind a protein, metabolize a compound,
15 confer toxicity to insects or pathogenic viruses, and the like) which can be selected for, and whereby the selected polynucleotide sequences are genetic sequences having a desired functionality and/or conferring a desired phenotypic property to an agricultural organism in which the polynucleotide has been transferred into. One method of identifying novel genetic sequences that possess a desired structure or
20 functional property in a plant or soil microbe, such as having an altered metabolism, involves the screening of a large library of recombinant sequences (which can be a component of a genome - e.g., part of a gene, non-coding transcriptional regulatory sequence, origin of replication, - or a complete genome of an organelle or microbe) for individual library members which possess the desired structure or functional
25 property conferred by the novel genetic sequence.

In a general aspect, the invention provides a method, termed "sequence shuffling" for use in plants and other agricultural organisms of interest such as fungi and even animals, for generating libraries of recombinant polynucleotides having a desired characteristic which can be selected or screened for in the relevant system,
30 e.g., in plant cell protoplasts or progeny thereof (plant cells, plants, etc.). Libraries of recombinant polynucleotides are generated from a population of related-sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. In the method, at

least two species of the related-sequence polynucleotides are combined in a recombination system suitable for generating sequence-recombined polynucleotides, wherein said sequence-recombined polynucleotides comprise a portion of at least one first species of a related-sequence polynucleotide with at least one adjacent portion of at least one second species of a related-sequence polynucleotide. Recombination systems suitable for generating sequence-recombined polynucleotides can be either: (1) in vitro systems for homologous recombination or sequence shuffling via amplification or other formats described herein, or (2) in vivo systems for homologous recombination or site-specific recombination as described herein, or template-switching of a retroviral genome replication event.

The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The selected sequence-recombined polynucleotides, which are typically related-sequence polynucleotides, can then be subjected to at least one recursive cycle wherein at least one selected sequence-recombined polynucleotide is combined with at least one distinct species of related-sequence polynucleotide (which may itself be a selected sequence-recombined polynucleotide) in a recombination system suitable for generating sequence-recombined polynucleotides, such that additional generations of sequence-recombined polynucleotide sequences are generated from the selected sequence-recombined polynucleotides obtained by the selection or screening method employed. In this manner, recursive sequence recombination generates library members which are sequence-recombined polynucleotides possessing desired characteristics. Such characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property.

Screening/selection produces a subpopulation of genetic sequences (or protoplasts, plants fungi or cells) expressing recombinant forms of gene(s) that have evolved toward acquisition of a desired property. These recombinant forms can then be subjected to further rounds of recombination and screening/selection in any order.

For example, a second round of screening/selection can be performed analogous to the first resulting in greater enrichment for genes having evolved toward acquisition of the desired property. Optionally, the stringency of selection can be increased between rounds (e.g., if selecting for drug resistance, the concentration of drug in the media can be increased). Further rounds of recombination can also be performed by an analogous strategy to the first round generating further recombinant forms of the gene(s) or genome(s). Alternatively, further rounds of recombination can be performed by any of the other molecular breeding formats discussed. Eventually, a recombinant form of the gene(s) or genome(s) is generated that has fully acquired the desired property.

The method of shuffling can generate libraries of polynucleotides (microbial enzymes adapted to perform a desired catalytic process in a plant cell, transgene polynucleotides) encoding selectable properties, which can compose all or a part of a genetic sequence or host cell transgene, wherein the library is suitable for function optimization of a gene or regulatory sequence or phenotypic screening. For example, the method can include (1) obtaining a first plurality of library members comprising an agricultural organism genome, gene, regulatory or replication sequence, or host cell transgene (or encoding sequence or expression cassette thereof), and obtaining from said library a polynucleotide, or copy thereof, complete or partial, of at least one selected library member having a detectable desired phenotype, optionally introducing mutations into said polynucleotide or copy(ies), and (2) shuffling these nucleic acids by any available method, e.g., by pooling and fragmenting, by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, typically producing random fragments or fragment equivalents, said selected polynucleotide(s) or copies to form fragments thereof under conditions suitable for PCR amplification, performing PCR amplification and optionally mutagenesis, and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides, whereby a substantial fraction (e.g., greater than about 10 percent) of the recombined polynucleotides of said shuffled pool are not present in the first plurality of selected library members, said shuffled pool composing a library of shuffled selected variant sequences or transgene sequences suitable for functional screening or phenotype screening. Optionally, the method comprises the additional step of screening the

library members of the shuffled pool to identify individual shuffled library members having the desired functional ability or phenotype. The novel shuffled genes, genome sequences, and transgene sequences that are identified from such libraries can be used and/or can be subjected to one or more additional cycles of shuffling and/or functional optimization or phenotype selection for further optimization. The method can be modified such that the step of selecting is for a phenotypic characteristic other than a metabolic trait, gene function, transcriptional regulatory sequence function, or the like. Oligonucleotide and in silico shuffling approaches can also be used.

In an embodiment, the first plurality of selected library members is fragmented and homologously recombined by PCR in vitro. Fragment generation is by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, such as described herein and in WO95/22625 published 24 August 1995, and in commonly owned U.S.S.N. 08/621,859 filed 25 March 1996, PCT/US96/05480 filed 18 April 1996, which are incorporated herein by reference). Stuttering is fragmentation by incomplete polymerase extension of templates. A recombination format based on very short PCR extension times can be employed to create partial PCR products, which continue to extend off a different template in the next (and subsequent) cycle(s), and effect de facto fragmentation. Template-switching and other formats which accomplish sequence shuffling between a plurality of sequence-related polynucleotides can be used. Such alternative formats will be apparent to those skilled in the art.

In an embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologously recombined to form shuffled library members in vivo.

In an embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologously recombined to form shuffled library members in vivo.

In an embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and

homologously recombined by intra-vector or inter-vector recombination to form shuffled library members in vivo.

In an embodiment, first plurality of selected library members is replicated under conditions wherein retroviral template switching between at least two xenogeneic genomes cloned into retrovirus vectors occurs, typically involving non-retroviral genes cloned into a retroviral replication system.

Other viral (and viral vector) systems such as gemini viruses, positive stranded RNA viruses and DNA viruses can be used.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity. The recombination cycles (in vitro or in vivo) can be performed in any order desired by the practitioner.

The present invention provides a method for generating libraries of shuffled polynucleotides suitable for functional screening (i.e., which is measured without respect to a phenotype conferred on a plant or related agricultural organism) or phenotypic screening (i.e., which is detected as a phenotype of a plant or other agricultural organism). The method generally comprises (1) obtaining a first plurality of selected library member polynucleotides comprising a polynucleotide conferring a selectable phenotype, and wherein said selected library member polynucleotides comprise a region of substantially identical sequence, optionally introducing mutations into said library member polynucleotides or copies, and (2) pooling and fragmenting, by chemical fragmentation, nuclease digestion, partial extension PCR amplification, PCR stuttering, site-specific recombination, or other suitable fragmenting means, typically producing random fragments or fragment equivalents, to form fragments thereof under conditions suitable for PCR amplification, performing PCR amplification and optionally mutagenesis, and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool are not present in the first plurality of selected library member polynucleotides, said shuffled pool composing a library of shuffled polynucleotide sequences ("shufflants") suitable for screening, either directly or subsequent to transformation into a host cell (e.g., a plant cell or microorganism). The method can be modified such that the step of selecting is for a phenotypic characteristic not naturally found in the host organism (e.g., for a herbicide catalytic

activity, viral resistance, drug resistance, or other non-native detectable phenotype conferred on a host cell or organism). Alternatively, the method can be modified such that the step of selecting is for a modified phenotype which is enhanced or diminished, or otherwise changed in character, as compared to the phenotype which naturally occurs in the host cell or host organism.

In one embodiment, the first plurality of selected library members is fragmented and homologously recombined by PCR in vitro. Fragment generation is by nuclease digestion, partial extension PCR amplification, PCR stuttering, or other suitable fragmenting means, such as described herein and in the documents incorporated herein by reference. Stuttering is fragmentation by incomplete polymerase extension of templates.

In one embodiment, the first plurality of selected library members is fragmented in vitro, the resultant fragments transferred into a host cell or organism and homologously recombined to form shuffled library members in vivo. In an aspect, the host cell is a plant cell which has been engineered to contain enhanced recombination systems, such as an enhanced system for general homologous recombination (e.g., a plant expressing a recA protein or a plant recombinase from a transgene or plant virus) or a site-specific recombination system (e.g., a cre/LOX or frt/FLP system encoded on a transgene or plant virus).

In one embodiment, the first plurality of selected library members is cloned or amplified on episomally replicable vectors, a multiplicity of said vectors is transferred into a cell and homologously recombined to form shuffled library members in vivo in a plant cell, fungal cell, algae cell, or bacterial cell. Other cell types may be used, if desired.

In one embodiment, the first plurality of selected library members is not fragmented, but is cloned or amplified on an episomally replicable vector as a direct repeat or indirect (or inverted) repeat, which each repeat comprising a distinct species of selected library member sequence, said vector is transferred into a cell and homologously recombined by intra-vector or inter-vector recombination to form shuffled library members in vivo in a plant cell, algae cell, or microorganism.

In an embodiment, combinations of in vitro and in vivo shuffling are provided to enhance combinatorial diversity.

Without reciting the various generalized formats of polynucleotide sequence shuffling and selection described previously or hereinbelow, which will be referred to herein by the shorthand “shuffling”, the present invention provides methods, compositions, and uses related to creating novel or improved plants, plant
5 cells, algal cells, soil microbes, plant pathogens, pharmaceuticals, commensal microbes, or other plant-related organisms having art-recognized importance to the agricultural, horticultural, and argonomic areas (collectively, “agricultural organisms”).

In an aspect, the invention provides a method for creating or altering a
10 phenotype of an agricultural organism by introducing a shuffled polynucleotide into said agricultural organism to generate a modified agricultural organism having a phenotype conferred by the introduced shuffled polynucleotide.

The invention also provides the modified agricultural organisms made by this method, and uses thereof. In a variation of the basic method, the method
15 comprises the further step of performing a selection or screening step on the modified agricultural organism to identify or quantitate a detectable phenotypic property. In various embodiments, such phenotypes can be, for example and not limitation, a herbicide-resistance trait, organ morphology, life-cycle modification (e.g., conversion of a short-day plant into a long-day plant, rapid fruit formation, delayed ripening,
20 suppressed seed formation), metabolic biosynthesis (e.g., carbon-fixation efficiency, lipid content, bulk protein composition, starch content, etc.), or any phenotype that the artisan skilled in agriculture, botany, plant sciences, plant pathology, biochemistry, nutrition, food processing, or horticulture would recognize as a detectable phenotype.

In an aspect, the invention provides a method for obtaining
25 polynucleotide sequences conferring a desired phenotype on an agricultural organism, the method comprising the steps of: (1) contacting or transforming a population of plant cells, algae cells, bacterial cells, fungal cells plant viruses, plants or explanted organs therefrom, with a first plurality of polynucleotide species having at least one region of substantial sequence identity to support shuffling to generate a first
30 transformed population, (2) selecting, from the first transformed population, a subpopulation having at least one desired phenotype, and recovering from the subpopulation a plurality of selected polynucleotide species, (3) recombining, by shuffling, said plurality of said selected polynucleotide species, thereby generating a

collection of shuffled polynucleotide species, and (4) contacting or transforming a population of plant cells, algae cells, bacterial cells, plant viruses, plants or explanted organs therefrom, with said collection of shuffled polynucleotide species to generate a second transformed population, and (5) selecting, from the second transformed population, at least one cell or organism having at least one desired phenotype. In a variation, at least one, preferably a plurality of, selected, shuffled polynucleotide(s) are recovered from the at least one cell or organism selected from the second population and having at least one desired phenotype; the selected, shuffled polynucleotide(s) are subjected to at least one subsequent round of shuffling (with each other, with related unshuffled sequences, with spike sequences, with mutagenic methods, or the like), transformation or contacting, and selection; this additional step can be repeated iteratively (with or without modification or variance in one or more cycles) from 1 to about 1000 cycles or as deemed suitable by the practitioner. Typically, the recombination in step (3) is performed in vitro or by an in vivo recombination method which substantially does not occur naturally in a plant cell at a recombination frequency of more than 10% of the frequency of the recombination methods described herein for polynucleotide sequence shuffling.

In certain variations, naturally occurring in vivo recombination mechanisms of plants, agricultural microorganisms, or vector-host cells for intermediate replication can be used in conjunction with a collection of shuffled polynucleotide sequence variants having a desired phenotypic property to be optimized further; in this way, a natural recombination mechanism can be combined with intelligent selection of variants in an iterative manner to produce optimized variants by "forced evolution", wherein the forced evolved variants are not expected to, nor are observed to, occur in nature, nor are predicted to occur at an appreciable frequency. The practitioner may further elect to supplement and/or the mutational drift by introducing intentionally mutated polynucleotide species suitable for shuffling, or portions thereof, into the pool of initial polynucleotide species and/or into the plurality of selected, shuffled polynucleotide species which are to be recombined. Mutational drift may also be supplemented by the use of mutagens (e.g., chemical mutagens or mutagenic irradiation), or by employing replication conditions which enhance the mutation rate.

The invention provides a method of performing recursive shuffling on a transgene portion or complete transgene, comprising: (1) introducing into a population of site-specific recombination plant cells a site-specific recombination transgene having loxP or FLP sites, or equivalents, and obtaining site-specific
5 integration or recombination of the transgene into a site-specific target site in the plant genome, (2) selecting from the population of plant cells a subpopulation having or encoding a desired phenotype, which may be an enzymatic function, a morphological trait observable following regeneration from the plant cell, or the like, (3) recovering a plurality of transgene sequences from the subpopulation, (4) shuffling the recovered
10 transgene sequences to create a shufflant library of transgenes having suitable site-specific recombination site(s), and (5) repeating steps 1 through 5 with the shufflant library of transgenes for at least one cycle of recursion, preferably for sufficient iterative cycles until the desired phenotype is evolved to the satisfaction of the practitioner. The invention provides the use of these site-specific recombination
15 system components (site-specific plant cells, site-specific transgene, and the like). In an embodiment, the selection step involves a biochemical assay or herbicide resistance assay that can be performed in plant cell culture without substantial development of an adult plant organism, and preferably is done in a high-throughput format, as by cell colony screening (e.g., using a reporter system in the cells) or by
20 multiwell plate format, or the like.

The invention provides regenerable plant cells and non-regenerable plant cell lines having homologous recombination systems with a detectable recombination frequency of at least 50 percent greater than the naturally-occurring plant cells of the same species and cell type. An embodiment comprises a plant cell
25 expressing a transgene-encoded heterologous recombinase (e.g., recA or the like), which may be of plant origin, animal origin (e.g., a general recombinase, the V-D-J recombinase, and the like), fungal origin, or bacterial origin (e.g., recA). A method of the invention employs such plant cells expressing a recombinase and homologous transgene constructs, to facilitate homologous gene targeting and homologous
30 transgene integration into plant genomes so as to either inactivate or replace an endogenous plant gene, and/or to homologously integrate a heterologous gene into a plant genome.

The invention also provides for the shuffled polynucleotide sequence(s) conferring the desired phenotype(s) on an agricultural organism, and the modified agricultural organisms themselves, produced by the method of polynucleotide sequence shuffling; the exact structures of said produced polynucleotide sequences and modified agricultural organisms are definable *a priori* only by reference to the method by which they are generated. Thus, the invention includes a shuffled polynucleotide sequence conferring the desired phenotype, or a plurality thereof, produced by the methods described herein. The shuffled polynucleotides(s) produced thereby are easily distinguishable from naturally occurring genome sequences by virtue of their atypical modified or novel phenotype(s) which is/are normally not present in the population of naturally occurring agricultural organism. The shuffled polynucleotide sequence can be further distinguished from naturally-occurring plant, animal, or microbe genome sequences by reference to sequence databases and published sequence data, wherein the shuffled polynucleotide will generally comprise a constellation of mutations as compared to the reference dataset which would be recognized by the skilled artisan as a polynucleotide sequence which is substantially improbable of having evolved by natural evolution or classical breeding.

In a variation of the basic method, one or more encoding sequences or transcriptional regulatory sequences derived from a plant genome are jointly or separately optimized (or improved for function) in a predetermined plant cell and/or host plant species as distinct genetic elements isolated from the remainder of the plant genome. The optimized or improved portions of the encoding sequence and/or transcriptional regulatory sequence is then introduced into the plant genome(s). In a variation, the optimized or improved portions can be used in conjunction with one or more heterologous polynucleotide sequence(s), such as genes or transcriptional regulatory sequences from other plant species or from non-plant genomes to confer a desired functional or structural property, such as transcriptional regulation or translational regulation, to the improved portions. Optimized or improved portions of a plant gene often can be marketed as a commercial product, either alone or in combination with one or more heterologous sequences.

The invention also encompasses compositions of such shuffled plant polynucleotides encoding at least one modified phenotype of an agricultural

organism. The compositions can include a plurality of species of shuffled polynucleotides, or can represent a single purified polynucleotide species. Certain shuffled polynucleotides encode variants which possess detectable phenotypes that are not naturally occurring and which can be selected for; selected phenotypes often
5 are characterized by desirable properties.

Additional Shuffling Formats-- Oligonucleotide mediated recombination and "In Silico" Recombination

In addition to the formats for shuffling noted above, at least two additional related formats are useful in the practice of the present invention, i.e., for
10 producing libraries of shuffled materials to be screened in protoplasts. These additional methods can be used individually or in combination with each other and with the formats noted herein, e.g., those above.

The first, referred to as "in silico" shuffling utilizes computer algorithms to perform "virtual" shuffling using genetic operators in a computer. As
15 applied to the present invention, gene sequence strings are recombined in a computer system and desirable products (such as libraries for transduction into protoplasts) are made, e.g., by reassembly PCR of synthetic oligonucleotides. In silico shuffling is described in detail in Selifonov and Stemmer in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING
20 DESIRED CHARACTERISTICS" filed 02/05/1999, USSN 60/118854. In brief, genetic operators (algorithms which represent given genetic events such as point mutations, recombination of two strands of homologous nucleic acids, etc.) are used to model recombinational or mutational events which can occur in one or more nucleic acid, e.g., by aligning nucleic acid sequence strings (using standard alignment
25 software, or by manual inspection and alignment) and predicting recombinational outcomes. The predicted recombinational outcomes are used to produce corresponding molecules, e.g., by oligonucleotide synthesis and reassembly PCR.

The second useful format is referred to as "oligonucleotide mediated shuffling" in which oligonucleotides corresponding to a family of related homologous
30 nucleic acids (e.g., as applied to the present invention, interspecific or allelic variants of a nucleic acid) which are recombined to produce selectable nucleic acids. This format is described in detail in Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed February 5, 1999, USSN 60/118,813

and Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" filed June 24, 1999, USSN 60/141,049. The technique can be used to recombine homologous or even non-homologous nucleic acid sequences.

5 One advantage of the oligonucleotide-mediated recombination is the ability to recombine homologous nucleic acids with low sequence similarity, or even non-homologous nucleic acids. In these low-homology oligonucleotide shuffling methods, one or more set of fragmented nucleic acids are recombined, e.g., with a set of crossover family diversity oligonucleotides. Each of these crossover oligonucleotides have a plurality of sequence diversity domains corresponding to a
10 plurality of sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity. The fragmented oligonucleotides, which are derived by comparison to one or more homologous or non-homologous nucleic acids, can hybridize to one or more region of the crossover oligos, facilitating recombination.

15 When recombining homologous nucleic acids, sets of overlapping family gene shuffling oligonucleotides (which are derived by comparison of homologous nucleic acids and synthesis of oligonucleotide fragments) are hybridized and elongated (e.g., by reassembly PCR), providing a population of recombined nucleic acids, which can be selected for a desired trait or property. Typically, the set
20 of overlapping family shuffling gene oligonucleotides include a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target nucleic acids.

Typically, family gene shuffling oligonucleotide are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence
25 identity and regions of sequence diversity. A plurality of family gene shuffling oligonucleotides are synthesized (serially or in parallel) which correspond to at least one region of sequence diversity.

Sets of fragments, or subsets of fragments used in oligonucleotide shuffling approaches can be provided by cleaving one or more homologous nucleic
30 acids (e.g., with a DNase), or, more commonly, by synthesizing a set of oligonucleotides corresponding to a plurality of regions of at least one nucleic acid (typically oligonucleotides corresponding to a full-length nucleic acid are provided as members of a set of nucleic acid fragments). In the shuffling procedures herein, these

cleavage fragments can be used in conjunction with family gene shuffling oligonucleotides, e.g., in one or more recombination reaction to produce recombinant nucleic acids.

Codon Modification Shuffling

5 In addition to the procedures noted above, libraries of codon-altered nucleic acids can be created to take advantage of non-naturally occurring sequence space. Procedures for codon modified shuffling are described in detail in SHUFFLING OF CODON ALTERED GENES, Phillip A. Patten and Willem P.C. Stemmer, filed September 29, 1998, USSN 60/102362 and in SHUFFLING OF
10 CODON ALTERED GENES, Phillip A. Patten and Willem P.C. Stemmer, filed January 29, USSN 60/117729. In brief, by synthesizing nucleic acids in which the codons which encode polypeptides are altered, it is possible to access a completely different mutational cloud upon subsequent mutation of the nucleic acid. This increases the sequence diversity of the starting nucleic acids for shuffling protocols,
15 which alters the rate and results of forced evolution procedures. Codon modification procedures can be used to modify any nucleic acid, e.g., prior to performing DNA shuffling, or codon modification approaches can be used in conjunction with Oligonucleotide Shuffling procedures as described supra.

In these methods, a first nucleic acid sequence encoding a first
20 polypeptide sequence is selected. A plurality of codon altered nucleic acid sequences, each of which encode the first polypeptide, or a modified or related polypeptide, is then selected (e.g., a library of codon altered nucleic acids can be selected in a biological assay which recognizes library components or activities), and the plurality of codon-altered nucleic acid sequences is recombined to produce a target codon
25 altered nucleic acid encoding a second protein. The target codon altered nucleic acid is then screened for a detectable functional or structural property, optionally including comparison to the properties of the first polypeptide and/or related polypeptides. The goal of such screening is to identify a polypeptide that has a structural or functional property equivalent or superior to the first polypeptide or related polypeptide. A
30 nucleic acid encoding such a polypeptide can be used in essentially any procedure desired, including introducing the target codon altered nucleic acid into a cell, vector, virus, attenuated virus (e.g., as a component of a vaccine or immunogenic composition), transgenic organism, or the like.

Phenotypic Selection

The present method can be used to create variant plant genes which exhibit altered function, stability, or expression by employing the rapid forced evolution of shuffling to generate variant genetic sequences that are adapted to the desired phenotype which is expressible in plant cell culture or in a regenerated plant or plant organ. The method is general and can be employed to modify a genetically conferred phenotype of substantially any agricultural organism suitable for recursive sequence shuffling.

The present method can also be employed to force evolution of plant host cells and polygenic transgenes to support enhanced transformation by *Agrobacterium* Ti plasmid or biolistics and/or minimize or reduce collateral genetic damage to the plant genome and progeny cells. By recursive shuffling and selection, it is possible to force the evolution of transgene-encoded proteins which permit facile transformation of substantially any regenerable plant cell. Multiple genetic sequences may be allowed to co-evolve, or the individual genetic sequences can be optimized individually and later recombined.

Although described with specificity with respect to higher plants, it is believed that the present method can be used with substantially any type of agricultural organism having a genome or gene portion suitable for *in vitro* or *in vivo* sequence shuffling with expression in plant cells and phenotype selection thereon. The recovered sequences can be shuffled with other genetic sequences and/or with one or more spiked polynucleotide specie(s) (e.g., mutation-bearing gene sequences or mutation-bearing sequences), which may include optimized components of a genotype that have been separately optimized by shuffling. Optimized components typically can include expression cassettes encoding plant or microbe metabolic genes, plant viral sequences, origins or replication, non-coding sequences important for replication, transcriptional control sequences, xenogeneic proteins, and the like. It is also possible to combine one or more cycle(s) of individual component/segment evolution with one or more cycle(s) of collective component/segment evolution, in any order.

In an aspect of the invention, a plurality of genetic sequences are shuffled and the resultant shuffled genetic sequences are selected for the capacity to

confer a desired phenotype to a host cell or organism harboring the shuffled sequence(s).

The present invention provides a method for generating libraries of genomes or genetic sequences suitable for phenotype screening, such as to generate enhanced function in a cell type and/or agricultural organism species, modify metabolism, resistance phenotype, or other desired property. The method comprises (1) obtaining a first plurality of library members comprising a genome polynucleotide or portion thereof, (2) pooling and fragmenting said polynucleotides or copies to form fragments thereof under conditions suitable for PCR amplification and thereby homologously recombining said fragments to form a shuffled pool of recombined polynucleotides comprising novel combinations of sequences, whereby a substantial fraction (e.g., greater than 10 percent) of the recombined polynucleotides of said shuffled pool comprise genome sequence combinations which are not present in the first plurality of library members, said shuffled pool composing a library of viral genome sequences comprising sequence combinations suitable for phenotype screening. Optionally, the plurality of selected shuffled library members can be shuffled and screened iteratively, from 1 to about 1000 cycles or as desired until library members having a desired binding affinity are obtained. Often, from 2 to 25 cycles of recursion are performed before a sufficiently optimized shufflant (i.e., selected shuffled library member) is obtained. The degree of optimization for any particular application will vary based on the specific intended use and other considerations (e.g., time, minimization of mutational drift, etc.) that are selected by the practitioner.

In general, the format of the assay used to select library members (e.g., protoplasts or reconstituted cells or organisms) will depend on the trait to be selected. For example, where the desired trait is herbicide resistance, survival of cells or protoplasts on media containing herbicides can be used to select desirable herbicide resistance traits. Similarly, where production of a metabolite (e.g., an oil, vitamin, phytohormone or phytochemical) is desired, the presence of the metabolite can be monitored, e.g., in a high-throughput fashion.

For example, one high throughput method for detecting analyte molecules from a complex biological mixture is by electrospray tandem mass spectrometry as taught in "HIGH THROUGHPUT MASS SPECTROMETRY" by

Sun Ai Raillard, USSN 60/119,766, filed 02/11/1999. In the '766 application, methods which utilize off-line parallel sample purification and fast flow-injection analysis, typically reducing the time of analysis to 30 to 40 seconds per sample. All steps starting from cell/protoplast picking, growth, sample preparation and analysis
5 are automated and can be carried out overnight by various robotic workstations.

The ability to detect a subtle increase in the performance of a shuffled library member over that of a parent strain relies on the sensitivity of the assay. The chance of finding the organisms having an improvement is increased by the number of individual mutants that can be screened by the assay. To increase the chances of
10 identifying a pool of sufficient size, a prescreen that increases the number of mutants processed by 10-fold can be used. The goal of the primary screen is to quickly identify mutants having roughly equal or better product titers than the parent strain(s) and to move only these mutants forward to liquid cell culture for subsequent analysis.

FORCED EVOLUTION OF GENES

15 The invention provides a means to evolve gene variants and/or host cells, as well as providing a model system for evaluating a library of agents to identify candidate agents that could find use as agricultural reagents (e.g., herbicide) for commercial applications.

The methods of the invention can be used to force the evolution of a
20 gene which has a beneficial property in one organism into a shuffled variant that can confer that same phenotype to a second organism in which the gene was substantially non-functional or inadequate.

Suitable transcriptional regulatory sequences include: cauliflower mosaic virus 19S and 35S promoters, NOS promoter, OCS promoter, rbcS promoter,
25 Brassica heat shock promoter, synthetic promoters, non-plant promoters modified, if advantageous, for function in plant cells, substantially any promoter that naturally occurs in a plant genome, promoters of plant viruses or Ti plasmids, tissue-preferential promoters or cis-acting elements, light-responsive promoters or cis-acting elements (e.g., rbcS LRE), hormone-responsive cis-acting elements, developmental
30 stage-specific promoters and cis-acting elements, viral promoters (e.g., from Tobacco Mosaic virus, Brome Mosaic Virus, Cauliflower Mosaic virus, and the like), and the like. In a variation, a transcriptional regulatory sequence from a first plant species is

optimized for functionality in a second plant species by application of recursive sequence shuffling.

Granularity of Shuffling

The "granularity" of a shuffling event refers to the relative average density of recombination joints per unit length (e.g., per kilobase) or per recombined polynucleotide molecule (e.g., per functional viral genome). For illustration, a coarse granularity could be an average of one or less recombination joint per polynucleotide resulting from a shuffling (i.e., sequence recombination event); a coarse granularity of shuffling generates a "low crossover library." It is often desirable to alter the granularity of shuffling in different recursion cycles, although this is not necessary in many cases. The granularity desired can frequently be selected by the practitioner and is typically accomplished by controlling the degree of recombination in the recombination format selected (e.g., for a fragmentation/reassembly format, a high degree of fragmentation will generate a small average fragment size and hence a finer granularity; increasing the number of polynucleotide species shuffled can also be used to obtain finer granularity, among other ways apparent to those skilled in the art upon review of the many references incorporated herein related to shuffling). The average size of segment from the parental sequence(s) represented in the library of sequence-recombined polynucleotides is denoted as the "average segment length", and may be expressed by unit length (e.g., per kilobase) or as a fraction of the parental sequence (e.g., one-quarter genome of HIV-1).

If a mutational strategy is employed, it is frequently desirable to select a granularity which results in an average segment length wherein, on average, one mutation (or slightly less) per segment is present.

The present method permits the construction of a library of shuffled genes (or gene portions) wherein the library contains a population of shuffled genes of any granularity desired by the practitioner. Libraries prepared from a plurality of parental genes can be made to have substantially any granularity; for example a gene library having, on average, at least two recombination joints (e.g., three distinct segments) per sequence-recombined genome can be generated, as can viral genomes having three, four, five, six, seven, eight, nine, ten, or more recombination joints (e.g., a genomic polynucleotide composed of 4, 5, 6, 7, 8, 9, 10, or 11 or more distinct sequence segments).

Spiking

The basic sequence shuffling methodology can be used to shuffle a collection of related sequences, wherein most or all of the related sequences substantially span a certain physical portion of a gene or genome (e.g., a structural gene, a transcriptional regulatory sequence, a replication origin, or an entire viral genome). For example, the collection of related polynucleotides could represent, e.g., alleles of a gene locus, variant genes). However, in some embodiments it is desirable to focus evolutionary pressure principally on one or more discrete segments of a genomic polynucleotide (e.g., a specific) or of a particular gene (e.g., on a specific functional domain or conserved sequence of a gene). One methodological modification to focus sequence diversity on a particular segment of a genome is to “spike” a recombination reaction with additional polynucleotides which represent only a subset of the locus being shuffled. These “spiking polynucleotides” can enhance the potential sequence diversity at the locus subset (e.g., randomly or pseudorandomly increase mutation density at the locus subset), or can overrepresent (or underrepresent) certain predetermined sequences in order to steer the sequence diversity in a predetermined direction (e.g., to overrepresent mutations which tend to produce a beneficial result based on prior results).

Backcrossing

After a desired phenotype is acquired to a satisfactory extent by a selected shuffled gene or portion thereof, it is often desirable to remove mutations which are not essential or substantially important to retention of the desired phenotype (“superfluous mutations”). Superfluous mutations can be removed by backcrossing, which is shuffling the selected shuffled gene(s) with one or more parental gene and/or naturally-occurring gene(s) (or portions thereof) and selecting the resultant collection of shufflants for those species that retain the desired phenotype. By employing this method, typically in two or more recursive cycles of shuffling against parental or naturally-occurring viral genome(s) (or portions thereof) and selection for retention of the desired phenotype, it is possible to generate and isolate selected shufflants which incorporate substantially only those mutations which confer the desired phenotype, whilst having the remainder of the genome (or portion thereof) consist of sequence which is substantially identical to the parental (or wild-type) sequence(s). As one example of backcrossing, a pea Rubisco subunit gene (small subunit) can be shuffled

and selected for the capacity to substantially function in any Angiosperm plant cells; the resultant selected shufflants can be backcrossed with one or more Rubisco genes of a particular plant species and selected for the capacity to retain the capacity to confer the phenotype. After several cycles of such backcrossing, the backcrossing
5 will yield gene(s) which contain the mutations necessary for the desired phenotype, and will otherwise have a genomic sequence substantially identical to the genome(s) of the host genome.

Isolated components (e.g., genes, regulatory sequences, packaging sequences, replication origins, and the like) can be optimized and then backcrossed
10 with parental sequences so as to obtain optimized components which are substantially free of superfluous mutations.

Transgenic Hosts

Transgenes and expression vectors can be constructed by any suitable method known in the art; by either PCR or RT-PCR amplification from a suitable cell
15 type or by ligating or amplifying a set of overlapping synthetic oligonucleotides; publicly available sequence databases and the literature can be used to select the polynucleotide sequence(s) to encode the specific protein desired, including any mutations, consensus sequence, or mutation kernel desired by the practitioner. The coding sequence(s) are operably linked to a transcriptional regulatory sequence and, if
20 desired, an origin of replication. Antisense or sense-suppression transgenes and genetic sequences can be optimized or adapted for particular host cells and organisms by the described methods.

The transgene(s) and/or expression vectors are transferred into host cells, protoplasts, pluripotent embryonic plant cells, microbes, or fungi by a suitable
25 method, such as for example lipofection, electroporation, microinjection, biolistics, Agrobacterium tumefaciens transduction of Ti plasmid, calcium phosphate precipitation, PEG-mediated DNA uptake, electroporation, electrofusion, or other method. Stable transfectant host cells can be prepared by art-known methods, as can transgenic cell lines.

Phenotypic Traits

30 A variety of such traits also include traits (or "phenotypic traits" or "phenotypes") are selectable with appropriate procedures and sufficient numbers of transgenes. Such traits include, but are not limited to, visible traits, environmental or

stress related traits, disease related traits and ripening traits, such traits also include flower or plant color, flower shape and size, leaf shape and size, flower number per plant, leaf number per plant, pest resistance, plant height, plant bushiness, time to flowering, cold hardiness, drought tolerance, tolerance to high temperatures, chemical resistance, flavor, and aroma. These traits are dependent upon the synthesis of structural proteins and enzymes which catalyze biosynthetic or degradative reactions of plant metabolism.

Target Plants

As used herein, "plant" refers to either a whole plant, a plant part, a plant cell, or a group of plant cells. The class of plants which can be used in the method of the invention is generally as broad as the class of higher plants amenable to protoplast transformation techniques, including both monocotyledonous and dicotyledonous plants. It includes plants of a variety of ploidy levels, including polyploid, diploid and haploid, and may employ non-regenerable cells for certain aspects which do not require development of an adult plant for selection or in vivo shuffling.

Transformation

The transformation of plants and protoplasts in accordance with the invention may be carried out in essentially any of the various ways known to those skilled in the art of plant molecular biology. See, in general, Methods in Enzymology Vol. 153 ("Recombinant DNA Part D") 1987, Wu and Grossman Eds., Academic Press, incorporated herein by reference. As used herein, the term transformation means alteration of the genotype of a host plant by the introduction of a nucleic acid sequence. The nucleic acid sequence need not necessarily originate from a different source, but it will, at some point, have been external to the cell into which it is to be introduced.

In one embodiment, the foreign nucleic acid is mechanically transferred by microinjection directly into plant cells by use of micropipettes. Alternatively, the foreign nucleic acid may be transferred into the plant cell by using polyethylene glycol. This forms a precipitation complex with the genetic material that is taken up by the cell (e.g., by incubation of protoplasts with "naked DNA" in the presence of polyethyleneglycol)(Paszkowski et al., (1984) EMBO J. 3:2717-22; Baker

et al (1985) Plant Genetics, 201-211; Li et al. (1990) Plant Molecular Biology Report 8(4)276-291].

In another embodiment of this invention, the introduced gene may be introduced into the plant or other cells by electroporation (Fromm et al., (1985)
5 "Expression of Genes Transferred into Monocot and Dicot Plant Cells by Electroporation," Proc. Natl Acad. Sci. USA 82:5824, which is incorporated herein by reference). In this technique, plant protoplasts are electroporated in the presence of plasmids or nucleic acids containing the relevant genetic construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the
10 introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form a plant callus. Selection of the transformed plant cells with the transformed gene can be accomplished using phenotypic markers.

Cauliflower mosaic virus (CaMV) may also be used as a vector for introducing the foreign nucleic acid into plant and other cells (Hohn et al., (1982)
15 "Molecular Biology of Plant Tumors," Academic Press, New York, pp.549-560; Howell, United States Patent No. 4,407,956). CaMV viral DNA genome is inserted into a parent bacterial plasmid creating a recombinant DNA molecule which can be propagated in bacteria. After cloning, the recombinant plasmid again may be cloned and further modified by introduction of the desired DNA sequence into the unique
20 restriction site of the linker. The modified viral portion of the recombinant plasmid is then excised from the parent bacterial plasmid, and used to inoculate the plant cells or plants. Similarly, tobacco mosaic virus, potato virus or other viral systems can be used.

Another method of introduction of nucleic acid segments is high
25 velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327:70-73). Although typically only a single introduction of a new nucleic acid segment is required, this method particularly provides for multiple introductions.

A method of introducing the nucleic acid segments into plant cells is to
30 infect a plant cell, an explant, a meristem or a seed with Agrobacterium tumefaciens transformed with the segment. Under appropriate conditions known in the art, the transformed plant cells are grown to form shoots, roots, and develop further into plants. The nucleic acid segments can be introduced into appropriate plant cells, for

example, by means of the Ti plasmid of Agrobacterium tumefaciens. The Ti plasmid is transmitted to plant cells upon infection by Agrobacterium tumefaciens, and is stably integrated into the plant genome (Horsch et al., (1984) "Inheritance of Functional Foreign Genes in Plants," Science, 233:496-498; Fraley et al., (1983) Proc. Natl. Acad. Sci. USA 80:4803).

5 Ti plasmids contain two regions essential for the production of transformed cells. One of these, named transfer DNA (T DNA), induces tumor formation. The other, termed virulent region, is essential for the introduction of the T DNA into plants. The transfer DNA region, which transfers to the plant genome, can be increased in size by the insertion of the foreign nucleic acid sequence without its transferring ability being affected. By removing the tumor-causing genes so that they no longer interfere, the modified Ti plasmid can then be used as a vector for the transfer of the gene constructs of the invention into an appropriate plant cell, such being a "disabled Ti vector."

15 All plant cells which can be transformed by Agrobacterium and whole plants regenerated from the transformed cells can also be transformed according to the invention so as to produce transformed whole plants which contain the transferred foreign nucleic acid sequence.

20 There are presently at least three different ways to transform plant cells with Agrobacterium:

(1) co-cultivation of Agrobacterium with cultured isolated protoplasts or plant cells, (2) transformation of cells or tissues with Agrobacterium, or (3) transformation of seeds, apices or meristems with Agrobacterium.

25 Method (1) uses, e.g., an established culture system that allows culturing protoplasts and plant regeneration from cultured protoplasts.

Method (2) uses, e.g., (a) plant cells or tissues that can be transformed by Agrobacterium and (b) induced to regenerate into whole plants.

Method (3) uses, e.g., micropropagation. In the binary system, to have infection, two plasmids are used: a T-DNA containing plasmid and a vir plasmid. Any one of a number of T-DNA containing plasmids can be used; the main caveat is that it may be desirable to be able to select independently for each of the two plasmids.

After transformation of the plant cell or plant, those plant cells or plants transformed by the Ti plasmid so that the desired DNA segment is integrated can be selected by an appropriate phenotypic marker. These phenotypic markers include, but are not limited to, antibiotic resistance, herbicide resistance or visual observation. Other phenotypic markers are known in the art and may be used in this invention.

PROTOPLAST TRANSFORMATION

Numerous protocols for establishment of transformable protoplasts from a variety of plant types and subsequent transformation of the cultured protoplasts are available in the art and are incorporated herein by general reference. For examples, see Hashimoto et al. (1990) Plant Physiol. 93: 857; Plant Protoplasts, Fowke LC and Constabel F, eds., CRC Press (1994); Saunders et al. (1993) Applications of Plant In Vitro Technology Symposium, UPM, 16-18 Nov. 1993; and Lyznik et al. (1991) BioTechniques 10: 295, each of which is incorporated herein by reference). Protoplast fusion is described by Shaffner et al., *Proc. Natl. Acad. Sci. USA* 77, 2163 (1980) and other exemplary procedures are described by Yoakum et al., US 4,608,339, Takahashi et al., US 4,677,066 and Sambrooke et al., at Ch. 16. Protoplast fusion has been reported between strains, species, and even diverse genera (e.g., yeast and chicken erythrocyte), as well as between plant protoplasts, fungal protoplasts and the like.

Protoplasts can be prepared for both bacterial and eukaryotic cells, including mammalian cells, fungal cells and plant cells, by several means, including chemical treatment to strip cell walls. For example, cell walls can be stripped by digestion with a cell wall degrading enzyme such as lysozyme in a 10-20% sucrose, 50 mM EDTA buffer. Conversion of cells to spherical protoplasts can be monitored by phase-contrast microscopy. Protoplasts can also be prepared by propagation of cells in media supplemented with an inhibitor of cell wall synthesis, or use of mutant strains lacking capacity for cell wall formation. Eukaryotic cells are optionally synchronized in G1 phase by arrest with inhibitors such as α -factor, *K. lactis* killer toxin, leflonamide and adenylate cyclase inhibitors.

Optionally, some protoplasts to be fused can be killed and/or have their DNA fragmented by treatment with ultraviolet irradiation, hydroxylamine or cupferon

(Reeves et al., *FEMS Microbiol. Lett.* 99, 193-198 (1992)). In this situation, killed protoplasts are referred to as donors, and viable protoplasts as acceptors. Using dead donors cells (e.g., comprising a previously introduced shuffled library) can be advantageous in subsequently recognizing fused cells with hybrid genomes. Further,
5 breaking up DNA in donor cells is advantageous for stimulating recombination with acceptor DNA. Optionally, acceptor and/or fused cells can also be briefly, but nonlethally, exposed to UV irradiation further to stimulate recombination in the protoplast or in protoplast fusions.

Once formed, protoplasts can be stabilized in a variety of osmolytes
10 and compounds such as sodium chloride, potassium chloride, sodium phosphate, potassium phosphate, sucrose, sorbitol, etc., e.g., in the presence of DTT. The combination of buffer, pH, reducing agent, and osmotic stabilizer can be optimized for different cell types. Protoplasts can be induced to fuse by treatment with a chemical such as PEG, calcium chloride or calcium propionate or electrofusion
15 (Tsoneva, *Acta Microbiologica Bulgaria* 24, 53-59 (1989)). A method of cell fusion employing electric fields has also been described. See Chang US, 4,970,154. Conditions can be optimized for different strains.

Fused cells are heterokaryons containing genomes from two or more component protoplasts. Fused cells can be enriched from unfused parental cells by
20 sucrose gradient sedimentation or cell sorting. The two nuclei in the heterokaryons can fuse (karyogamy) and homologous recombination can occur between the genomes. The chromosomes can also segregate asymmetrically resulting in regenerated protoplasts that have lost or gained whole chromosomes. The frequency of recombination can be increased by treatment with ultraviolet irradiation or by use
25 of strains overexpressing recA or other recombination genes, or the yeast rad genes, and cognate variants thereof in other species, or by the inhibition of gene products of *MutS*, *MutL*, or *MutD*. Overexpression can be either the result of introduction of exogenous recombination genes or the result of selecting strains, which as a result of natural variation or induced mutation, overexpress endogenous recombination genes.
30 The fused protoplasts are propagated under conditions allowing regeneration of cell walls, recombination and segregation of recombinant genomes into progeny cells from the heterokaryon and expression of recombinant genes. This process can be reiteratively repeated to increase the diversity of any set of protoplasts or cells. After,

or occasionally before or during, recovery of fused cells, the cells are screened or selected for evolution toward a desired property.

Subsequent rounds of recombination can be performed by preparing protoplasts from cells (or whole organisms, or protoplasts, depending on the format) surviving selection/screening in a previous round. The protoplasts are optionally fused, with recombination occurring in fused protoplasts. Cells, tissues or whole organisms are optionally regenerated from the fused protoplasts. This process can again be reiteratively repeated to increase the diversity of the starting population. Protoplasts, or regenerated or regenerating cells are subject to further selection or screening.

For additional details on whole cell/protoplast recombination methods, see, e.g., EVOLUTION OF WHOLE CELLS & ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION filed 07/15/1999, application No: PCT/US99/15972. In the methods of the '972 application, a variety of approaches for poolwise recombination of entire genomes are provided.

All plants for which corresponding protoplasts can be isolated and cultured can be transformed by the present invention so that whole plants are recovered which contain the transferred foreign gene. These cells can then be cultured into transgenic plants.

Suitable plants for protoplasting include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciohorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Further, it is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major cereal crop species, sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Limited knowledge presently exists on whether all of these plants can be transformed by Agrobacterium. Species which are a natural plant host for Agrobacterium may be transformable in vitro. Although monocotyledonous plants, and in particular, cereals

and grasses, are not natural hosts to Agrobacterium, work to transform them using Agrobacterium has also been successfully carried out by numerous investigators (Hooykas-Van Slogteren et al., (1984) Nature 311:763-764; Hernalsteens et al., (1984) EMBO J. 3:3039-41; Byteiber, et al. (1987) Proc. Natl. Acad. Sci. USA: 5345-5349; Graves and Goldman, (1986) Plant Mol. Biol 7: 43-50; Grimsley et al. (1988) Biochemistry 6: 185-189; WO 86/03776; Shimamoto et al. Nature (1989) 338: 274-276). Monocots may also be transformed by techniques or with vectors other than Agrobacterium. For example, monocots have been transformed by electroporation (Fromm et al. [1986] Nature 319:791-793; Rhodes et al. Science [1988] 240: 204-207), direct gene transfer (Baker et al. [1985] Plant Genetics 201-211), by using pollen-mediated vectors (EP 0 270 356), and by injection of DNA into floral tillers (de la Pena et al. [1987], Nature 325:274-276). Additional plant genera that may be transformed by Agrobacterium include Chrysanthemum, Dianthus, Gerbera, Euphorbia, Pelargonium, Ipomoea, Passiflora, Cyclamen, Malus, Prunus, Rosa, Rubus, Populus, Santalum, Allium, Lilium, Narcissus, Ananas, Arachis, Phaseolus and Pisum.

Important commercial crops that can be used in the methods of the invention include both monocots and dicots. Monocots include plants in the grass family (*Gramineae*), such as plants in the sub families *Fetuoideae* and *Poacoideae*, which together include several hundred genera including plants in the genera *Agrostis*, *Phleum*, *Dactylis*, *Sorgum*, *Setaria*, *Zea* (e.g., corn), *Oryza* (e.g., rice), *Triticum* (e.g., wheat), *Secale* (e.g., rye), *Avena* (e.g., oats), *Hordeum* (e.g., barley), *Saccharum*, *Poa*, *Festuca*, *Stenotaphrum*, *Cynodon*, *Coix*, the *Olyreae*, *Phareae* and many others. Plants in the family *Gramineae* are an example preferred target for the methods of the invention. Additional preferred targets include other commercially important crops, e.g., from the families *Compositae* (the largest family of vascular plants, including at least 1,000 genera, including important commercial crops such as sunflower), and *Leguminosae* or "pea family," which includes several hundred genera, including many commercially valuable crops such as pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, soybean, clover, alfalfa, lupine, vetch, *lotus*, sweet clover, wisteria, and sweetpea. Common crops applicable to the methods of the invention include *Zea mays* (corn), rice, soybean, sorghum, wheat, oats, barley, millet, sunflower, and canola.

Shuffling Fungi

In addition to plants, fungal cells can also be protoplasted and shuffled in the manner described herein for plants. Spores from a frozen stock, a lyophilized stock, or fresh from an agar plate are used to inoculate suitable liquid medium.

5 Spores are germinated resulting in hyphal growth. Mycelia are harvested, and washed by filtration and/or centrifugation. Optionally the sample is pretreated with DTT to enhance protoplast formation. Protoplasting is performed in an osmotically stabilizing medium (e.g., 1 M NaCl/20mM MgSO₄, pH 5.8) by the addition of cell wall-degrading enzyme (e.g., Novozyme 234). Cell wall degrading enzyme is removed by
10 repeated washing with osmotically stabilizing solution. Protoplasts can be separated from mycelia, debris and spores by filtration through miracloth, and density centrifugation. Protoplasts are harvested by centrifugation and resuspended to the appropriate concentration. This step may lead to some protoplast fusion. Fusion can be stimulated by addition of PEG (e.g., PEG 3350), and/or repeated centrifugation and
15 resuspension with or without PEG. Electrofusion can also be performed. Fused protoplasts can optionally be enriched from unfused protoplasts by sucrose gradient sedimentation (or other methods of screening described above). Fused protoplasts can optionally be treated with ultraviolet irradiation to stimulate recombination. Protoplasts are cultured on osmotically stabilized agar plates to regenerate cell walls
20 and form mycelia. The mycelia are used to generate spores, which are used as the starting material in the next round of shuffling.

Selection for a desired property can be performed either on regenerated mycelia or spores derived therefrom.

In an alternative method, protoplasts are formed by inhibition of one or
25 more enzymes required for cell wall synthesis. The inhibitor should be fungistatic rather than fungicidal under the conditions of use. Examples of inhibitors include antifungal compounds described by (e.g., Georgopapadakou & Walsh, *Antimicrob. Ag. Chemother.* 40, 279-291 (1996); Lyman & Walsh, *Drugs* 44, 9-35 (1992)). Other examples include chitin synthase inhibitors (polyoxin or nikkomycin compounds)
30 and/or glucan synthase inhibitors (e.g. echinocandins, papulocandins, pneumocandins). Inhibitors should be applied in osmotically stabilized medium. Cells stripped of their cell walls can be fused or otherwise employed as donors or hosts in genetic transformation/strain development programs.

Fungi which can be shuffled include filamentous fungi, which are particularly suited to performing the shuffling methods described above. Filamentous fungi are divided into four main classifications based on their structures for sexual reproduction: *Phycomycetes*, *Ascomycetes*, *Basidiomycetes* and the *Fungi Imperfecti*. *Phycomycetes* (e.g., *Rhizopus*, *Mucor*) form sexual spores in sporangium. The spores can be uni or multinucleate and often lack septated hyphae (coenocytic). *Ascomycetes* (e.g., *Aspergillus*, *Neurospora*, *Penicillium*) produce sexual spores in an ascus as a result of meiotic division. Asci typically contain 4 meiotic products, but some contain 8 as a result of additional mitotic division. *Basidiomycetes* include mushrooms, and smuts and form sexual spores on the surface of a basidium. In *holobasidiomycetes*, such as mushrooms, the basidium is undivided. In *hemibasidiomycetes*, such as rusts (*Uredinales*) and smut fungi (*Ustilaginales*), the basidium is divided. *Fungi imperfecti*, which include most human pathogens, have no known sexual stage.

Regeneration

Normally, regeneration will be involved in obtaining a whole plant or other organism from the transformation process. The term "transgenote" refers to the immediate product of the transformation process and to resultant whole transgenic plants.

The term "regeneration" as used herein, means growing a whole plant from a plant cell, a group of plant cells, a plant part or a plant piece (e.g. from a protoplast, callus, or tissue part).

Plant regeneration from cultural protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," Handbook of Plant Cell Cultures 1:124-176 (MacMillan Publishing Co. New York 1983); M.R. Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," Protoplasts, (1983) - Lecture Proceedings, pp.12-29, (Birkhauser, Basel 1983); P.J. Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," Protoplasts (1983) - Lecture Proceedings, pp. 31-41, (Birkhauser, Basel 1983); and H. Binding, "Regeneration of Plants," Plant Protoplasts, pp.21-73, (CRC Press, Boca Raton 1985). Other references relevant to protoplasting include include Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY (Payne); and Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture: Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg

New York) (Gamborg). Additional information is also found in commercial literature such as the Life Science Research Cell Culture catalogue (1998) from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-LSRCCC) and, e.g., the Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, MO) (Sigma-PCCS).

5 Regeneration from protoplasts varies from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the exogenous sequence is first made. In certain species embryo formation can then be induced from the protoplast suspension, to the stage of ripening and germination as natural embryos. The culture media will generally contain various amino acids and
10 hormones, such as auxin and cytokinins. It is sometimes advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and
15 repeatable.

Regeneration also occurs from plant callus, explants, organs or parts. Transformation can be performed in the context of organ or plant part regeneration. See, Methods in Enzymology, *supra*; also Methods in Enzymology, Vol. 118; and Klee et al., (1987) Annual Review of Plant Physiology, 38:467-486.

20 In vegetatively propagated crops, the mature transgenic plants are propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants for trialling, such as testing for production characteristics. Selection of desirable transgenes is made and new varieties are obtained thereby, and propagated vegetatively for commercial sale.

25 In seed propagated crops, the mature transgenic plants are self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the gene for the newly introduced foreign gene activity level. These seeds can be grown to produce plants that would produce the selected phenotype.

30 The inbreds according to this invention can be used to develop new hybrids. In this method a selected inbred line is crossed with another inbred line to produce the hybrid. The offspring resulting from the first experimental crossing of two parents is known in the art as the F1 hybrid, or first filial generation. Of the two

parents crossed to produce F1 progeny according to the present invention, one or both parents can be transgenic plants.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are covered by the invention, provided that these parts comprise cells which have been so transformed. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention, provided that these parts comprise the introduced DNA sequences. Progeny and variants, and mutants of the regenerated plants are also included within the scope of this invention.

Microspore Manipulation

Microspores are haploid (1n) male spores that develop into pollen grains. Anthers contain a large numbers of microspores in early-uninucleate to first-mitosis stages. Microspores have been successfully induced to develop into plants for most species, such as, e.g., rice (Chen, CC 1977 In Vitro. 13: 484-489), tobacco (Atanassov, I. et al. 1998 Plant Mol Biol. 38:1169-1178), Tradescantia (Savage JRK and Papworth DG. 1998 Mutat Res. 422:313-322), Arabidopsis (Park SK et al. 1998 Development. 125:3789-3799), sugar beet (Majewska-Sawka A and Rodrigues-Garcia MI 1996 J Cell Sci. 109:859-866), Barley (Olsen FL 1991 Hereditas 115:255-266) and oilseed rape (Boutillier KA et al. 1994 Plant Mol Biol. 26:1711-1723).

The plants derived from microspores are predominantly haploid or diploid (infrequently polyploid and aneuploid). The diploid plants are homozygous and fertile and can be generated in a relatively short time. Microspores obtained from F1 hybrid plants represent great diversity, thus being an excellent model for targeting and studying recombination. In addition, microspores can be transformed with T-DNA introduced by agrobacterium or other available means and then regenerated into individual plants. Furthermore, protoplasts can be made from microspores and they can be fused similar to what occur in fungi and bacteria.

Microspores, due to their complex ploidy and regenerating ability, provide a tool for plant whole genome shuffling. For example, if pollens from 4 parents are collected and pooled, and then used to randomly pollinate the parents, the progenies should have $2^4 = 16$ possible combinations. Assuming this plant has 7 chromosomes, microspores collected from the 16 progenies will represent $2^7 \times 16 =$

2048 possible chromosomal combinations. This number is even greater if meiotic processes occur. When diploid, homozygous embryos are generated from these microspores, in many cases, they are screened for desired phenotypes, such as herbicide- or disease- resistant. In addition, for plant oil composition these embryos
5 can be dissected into two halves: one for analysis the other for regeneration into a viable plant.

Protoplasts generated from microspores (especially the haploid ones) are pooled and fused. Microspores obtained from plants generated by protoplast fusion are optionally pooled and fused again, increasing the genetic diversity of the
10 resulting microspores.

Microspores can also be subjected to mutagenesis in various ways, such as by chemical mutagenesis, radiation-induced mutagenesis and, e.g., t-DNA transformation, prior to fusion or regeneration. New mutations which are generated can be recombined through the recursive processes described above and herein.

15 Vectors

Selection of an appropriate vector is relatively simple, as the constraints are minimal. The minimal traits of the vector are that the desired nucleic acid sequence be introduced in a relatively intact state. Thus, any vector which will produce a plant carrying the introduced DNA sequence should be sufficient. Also,
20 any vector which will introduce a substantially intact RNA which can ultimately be converted into a stably maintained DNA sequence should be acceptable.

Even a naked piece of DNA would be expected to be able to confer the properties of this invention, though at low efficiency. The decision as to whether to use a vector, or which vector to use, will be guided by the method of transformation
25 selected.

If naked nucleic acid introduction methods are chosen, then the vector need be no more than the minimal nucleic acid sequences necessary to confer the desired traits, without the need for additional other sequences. Thus, the possible vectors include the Ti plasmid vectors, shuttle vectors designed merely to maximally
30 yield high numbers of copies, episomal vectors containing minimal sequences necessary for ultimate replication once transformation has occurred, transposon vectors, homologous recombination vectors, mini-chromosome vectors, and viral vectors, including the possibility of RNA forms of the gene sequences. The selection

of vectors and methods to construct them are commonly known to persons of ordinary skill in the art and are described in general technical references (Methods in Enzymology, supra).

However, any additional attached vector sequences which will confer resistance to degradation of the nucleic acid fragment to be introduced, which assists in the process of genomic integration or provides a means to easily select for those cells or plants which are actually, in fact, transformed are advantageous and greatly decrease the difficulty of selecting useable transgenotes.

Recovery of Selected Polynucleotide Sequences

A variety of selection and screening methods will be apparent to those skilled in the art, and will depend upon the particular phenotypic properties that are desired. The selected shuffled genetic sequences can be recovered for further shuffling or for direct use by any applicable method, including but not limited to: recovery of DNA, RNA, or cDNA from cells (or PCR-amplified copies thereof) from cells or medium, recovery of sequences from host chromosomal DNA or PCR-amplified copies thereof, recovery of episome (e.g., expression vector) such as a plasmid, cosmid, viral vector, artificial chromosome, and the like, or other suitable recovery method known in the art. Libraries of nucleic acids are also thus obtained from populations of organisms, e.g., cells or protoplasts comprising shuffled nucleic acids. These secondary libraries can be used to transform additional protoplasts, plants, or the like.

Any suitable art-known method, including RT-PCR or PCR, can be used to obtain the selected shufflant sequence(s) for subsequent manipulation and shuffling.

The following example is given to illustrate the invention, but are not to be limiting thereof.

EXPERIMENTAL EXAMPLE

EXAMPLE 1: Selection of EPSP-synthase gene for glyphosate resistance in tobacco protoplasts

EPSP synthase (EPSPS) genes are isolated from commercially available cDNA libraries of Arabidopsis, tomato, tobacco, maize and other plants. The gene is alternatively isolated from cDNA prepared from poly (A+) mRNA from

floral organs of different parts (Gasser et al. J. Biol Chem. 263: 4280-4289, 1988, incorporated herein by reference). Primers for isolation of cDNA specific for EPSPS are designed based on consensus sequences derived from public information (J. Biol. Chem, above and Padgett et al. 1996 in Herbicide Resistant Crops Duke S (ed) pp 53-84) and used for gene isolation as described in the above citations. The EPSPS genes isolated from cDNAs of different plants contain the transit sequences for targeting of the genes to the chloroplasts.

The EPSPS genes from various plants, which have nucleotide homology in the range 75-93%, are shuffled according to published procedures for polynucleotide shuffling. Briefly, this procedure involves random fragmentation of the genes with DNase I and selecting nucleotide fragments of 100-300 bp. The fragments are reassembled based on sequence similarity by primerless PCR. Recombination as well as variable levels of mutations that are introduced by the PCR reaction generate the diversity. The assembled gene is cloned into a plasmid such as the Ti-based vector pBin19 used in *Agrobacterium tumefaciens*-mediated transformation. The schematic representation of the plasmid is shown in Figure 1 (see, Dyer WE in Herbicide Resistant Crops Duke S (ed.) pp 37-51). Shuffled EPSPS genes are cloned into multiple cloning sites shown in the plasmid and directly electroporated into tobacco protoplasts. Preparation of protoplasts from tobacco leaves and subsequence transformation and culturing conditions are described in the literature.

Transformed tobacco protoplasts, carrying EPSPS resistant to glyphosate are selected directly on a growth medium containing glyphosate. The level of glyphosate used is determined by plating untransformed tobacco protoplasts in a range of herbicide concentrations. At least 10x the lethal concentration (between 0.5 and 5 mM) is used for initial selection of glyphosate resistant lines. Transformed tobacco protoplasts are plated in the selection media. Those protoplasts containing the resistant gene grow into individual microcalli. EPSPS genes are isolated from this callus (or calli if multiple individuals are selected) and used for a subsequent rounds of sequence-shuffling and phenotype selection for glyphosate resistance. Eventually, the optimized gene is assayed for magnitude of resistance and quantification of other properties.

The resultant genetic sequence encoding glyphosate resistance is cloned into a plant cell protoplast capable of regeneration as a transgene or other stable, replication sequence that segregates with germplasm, an adult plant is regenerated, and the resultant regenerated plant species is bred to establish a
5 germplasm which can be used to produce glyphosate-resistance plants which can be sold commercially as seed or as vegetative plants.

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed,
10 and many modifications and variations are possible in light of the above teaching.

Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention.

All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was
15 specifically and individually indicated to be incorporated by reference for all purposes.

WHAT IS CLAIMED IS

1. A composition comprising a population of protoplast library members, wherein said protoplast library members each comprise a plant or fungal cell protoplast harboring intracellularly at least one species of a library of
5 heterologous polynucleotide sequences, each of said heterologous polynucleotide sequences operably linked to an expression sequence, or, if the heterologous polynucleotide sequence is a transcriptional regulatory sequence, operably linked to a reporter gene sequence.
2. The composition of claim 1, wherein the library of heterologous
10 polynucleotide sequences comprise at least 10 species of distinct heterologous polynucleotide sequences which share at least 70 percent sequence identity.
3. The composition of claim 1, wherein the library of heterologous polynucleotide sequences are substantially identical to a naturally-occurring gene
15 sequence in the genome of a species of plant, fungus, algae, dinoflagellate, bacterium, archaeobacterium, cyanobacterium, or plant pathogen, which naturally-occurring gene sequence is substantially or completely absent in the genome of the plant species from which the plant cell protoplasts harboring the library was produced.
4. The composition of claim 1, wherein the protoplast library members contain heterologous polynucleotides which are sequence-shuffled variants
20 of at least two parental polynucleotide species.
5. The composition of claim 1, wherein the protoplast library members comprise a heterologous nucleic acid encoding recombinase activity, or wherein the protoplast library members comprise a heterologous recombinase.
6. The composition of claim 5, wherein the heterologous
25 recombinase is selected from a bacterial RecA recombinase, an FLP recombinase or a Cre recombinase.

7. The composition of claim 1, wherein the protoplast library members are derived from mutant cells which express elevated levels of recombinase or mutator activity.

8. A library of protoplast, plant cell, fungal cell, plant or fungal library members, wherein said library members each comprise a cell or a protoplast harboring intracellularly at least one species of a selected shuffled library of heterologous polynucleotide sequences, produced by the steps of:
- (i.) transducing a first population of protoplasts with a first shuffled library population to produce a first transduced protoplast library;
 - 10 (ii.) selecting the transduced protoplast library, or a derivative thereof, for a desired activity; and,
 - (iii.) recombining selected nucleic acids from selected protoplast library members to produce the selected shuffled library;

wherein each of said heterologous polynucleotide sequences is operably linked to an expression sequence, or, if the heterologous polynucleotide sequence is a transcriptional regulatory sequence, operably linked to a reporter gene sequence, wherein the heterologous nucleic acids present in the shuffled library are homologous.

9. The library of claim 8, wherein the library of shuffled polynucleotide sequences comprise at least 10 species of distinct heterologous polynucleotide sequences which share at least 70 percent sequence identity.

10. The library of claim 8, wherein the selected nucleic acids are recombined between protoplasts by isolating the nucleic acids from the protoplasts and recombining the selected nucleic acids.

11. The library of claim 8, wherein the selected nucleic acids are recombined between protoplasts by fusing the protoplasts and permitting recombination to occur in the protoplasts.

12. The library of claim 8, wherein a derivative protoplast library is screened in step (ii), where the derivative library is produced recombining nucleic acids present in the first transduced protoplast library prior to selection.

13. The library of claim 8, wherein a derivative plant cell or organism library is screened in step (ii), wherein the plant cell or organism library is derived from the first protoplast library by a method comprising reconstituting the protoplast members of the library, or clonal or recombinational descendents thereof, into plant
5 cells.

14. A method for obtaining a desired polynucleotide sequence, comprising: selecting, from a population of protoplast library members or their clonal progeny, wherein said protoplast library members each comprise a plant cell protoplast harboring intracellularly one or a subset of a library of heterologous
10 polynucleotide sequences, a subpopulation of said library members which express a predetermined phenotype.

15. The method of claim 14, wherein the clonal progeny are selected from plant cells, fungal cells, plants and fungi.

16. The method of claim 14, further comprising recombining the
15 subset of heterologous library members prior to said selecting step.

17. The method of claim 14, further comprising making the protoplast library members by transducing a population of protoplasts with a shuffled nucleic acid library of sequences.

18. The method of claim 14, wherein the step of selecting comprises
20 assaying a detectable biochemical phenotype in library members and segregating into a subpopulation those library members which exhibit said detectable biochemical phenotype.

19. The method of claim 14, comprising the further step of recovering the heterologous polynucleotide sequences from said subpopulation of said library
25 members which express a predetermined phenotype thereby providing a collection of selected polynucleotide sequences, sequence-shuffling said selected polynucleotide sequences and performing at least one round(s) of transformation and selection for the desired phenotype.

20. The method of claim 19, further comprising reiteratively recombining the heterologous polynucleotide sequences prior to selection of the desired phenotype.

5 21. A method for rapid evolution of polynucleotide sequences conferring a desired or predetermined phenotype to at least one plant species, fungal species, algal species, or cyanobacterium, the method comprising:

- 10 (i) transferring a first population of sequence-shuffled polynucleotides comprising a genetic sequence into a plurality of plant or fungal cells or protoplasts to produce a first population of transformed plant or fungal cells or protoplasts, wherein the sequence-shuffled polynucleotides are expressible;
- 15 (ii) selecting, from the first population of transformed plant or fungal cells or protoplasts, and optionally from clonal progeny thereof, a plurality of genotypes present in said first population of transformed plant cells and expressing the desired phenotype, thereby generating a collection of selected genotypes;
- 20 (iii) producing a second population of sequence-shuffled polynucleotides comprising said genetic sequence obtained from the collection of selected genotypes and transferring said second population into a plurality of plant or fungal cells or protoplasts, thereby forming a second population of transformed plant or fungal cells or protoplasts, and optionally clonal progeny thereof; and,
- 25 (iv) selecting or identifying from the second population of transformed plant cells at least one genotype present in said second population of transformed plant cells and expressing the desired phenotype, thereby identifying at least one genotype comprising an evolved shuffled genetic sequence.

30 22. The method of claim 21, further comprising recombining the population of selected genotypes or the second population of sequence-shuffled polynucleotides prior to performing step iv.

23. The method of claim 21, wherein steps ii, iii, and iv are repeated iteratively until at least one genetic sequence possesses a satisfactory capacity to produce the desired phenotype.

5 24. The method of claim 23, wherein from 2 to 50 cycles of iterative shuffling, transfer into host cells, and selection are performed.

25. The method of claim 21, comprising the further step of transferring, into the germplasm of a plant species, the evolved shuffled genetic sequence encoding the genotype.

10 26. A method for identifying polynucleotide sequences encoding a predetermined phenotype for a plant cell, the method comprising:
(i) transforming a plurality of species of sequence-shuffled polynucleotides into protoplasts of plant cells which are clonal progeny of a predetermined non-regenerating plant cell line; and,
(ii) selecting transformed non-regenerable protoplasts or their clonal
15 progeny by segregating individual transformants or pools thereof which express a predetermined phenotype and recovering at least one polynucleotide sequence of a sequence-shuffled polynucleotide.

27. The method of claim 26, comprising the further step of culturing the transformed protoplasts on a semisolid medium in growth conditions to form a
20 population of microcalli, wherein substantially each microcallus comprises the clonal progeny of a transformed protoplast and subjecting the microcalli or portions thereof to selection for the desired phenotype(s).

28. The method of claim 26, wherein the sequence-shuffled polynucleotides comprise a selectable marker gene and the semisolid medium or
25 growth conditions initially select for transformants expressing the selectable marker gene which are capable of growth into microcalli whereas untransformed protoplasts and their progeny are substantially incapable of growth into microcalli.

29. The method of claim 26, wherein the transformed protoplasts are propagated as suspensions of callus cells wherein the clonal progeny of individual transformants are propagated in discrete culture vessels.

5 30. The method of claim 29, wherein the discrete culture vessels are wells of a multiwell culture plate.

31. A plant cell protoplast or clonal progeny thereof containing a sequence-shuffled polynucleotide which is not encoded by the naturally occurring genome of the plant cell protoplast.

10 32. The clonal progeny of claim 31, wherein the clonal progeny is a plant.

33. A collection of plant cell protoplasts transformed with a library of sequence-shuffled polynucleotides in expressible form.

15 34. A regenerated plant containing at least one species of replicable or integrated polynucleotide comprising a sequence-shuffled polynucleotide sequence in expressible form.

35. A kit for obtaining a polynucleotide encoding a predetermined phenotype, the kit comprising a plant cell line suitable for forming transformable protoplasts and a collection sequence-shuffled polynucleotides formed by in vitro sequence shuffling.

1/1

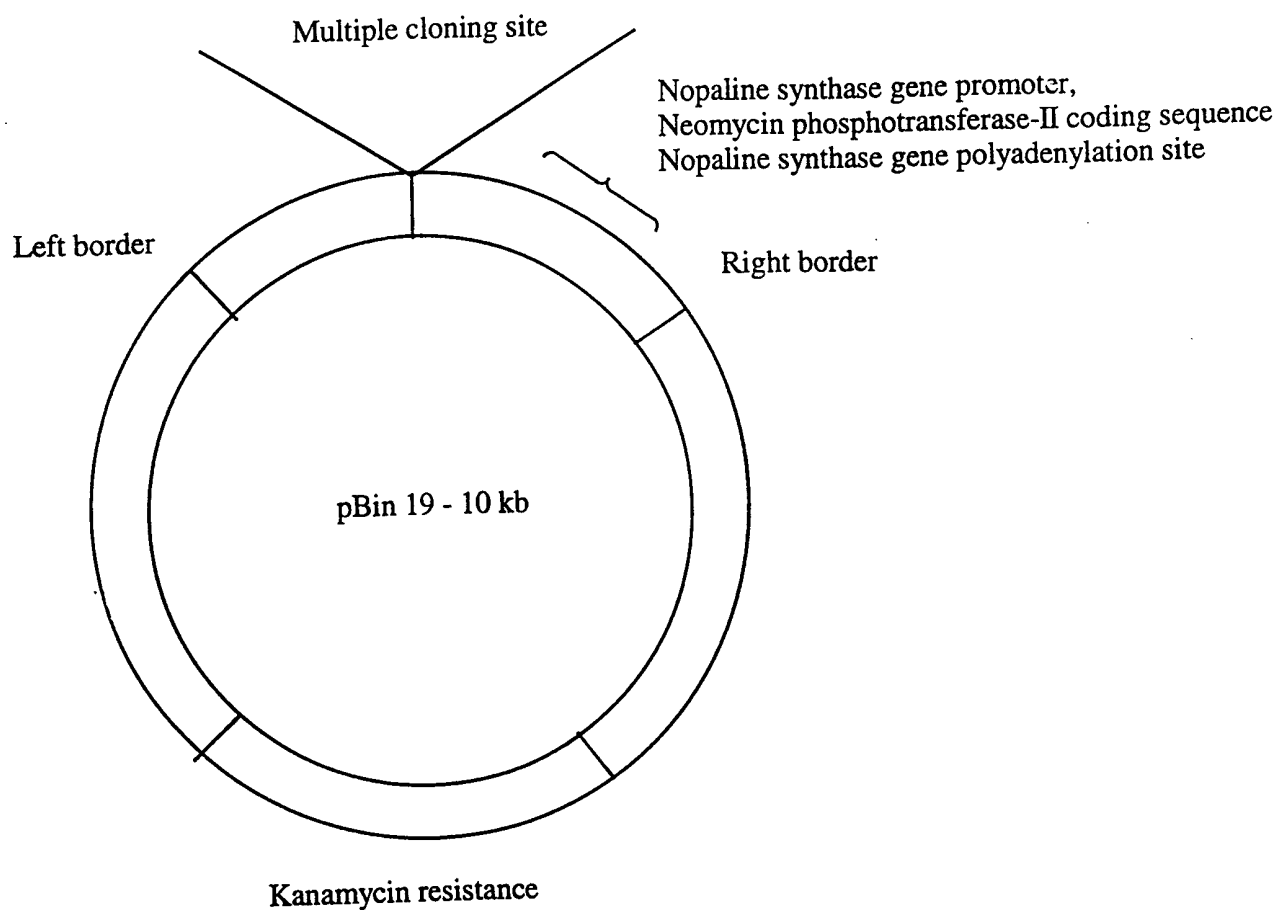


Figure 1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/19732

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 5/04, 15/29, 15/82; A01H 4/00
US CL : 435/69.1, 468, 453, 454; 800/278, 295,
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 468, 453, 454; 800/278, 295

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST, EAST, AGRICOLA, BIOSIS, MEDLINE, CAPLUS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	LYZNIK et al. Activity of Yeast FLP Recombinase in Maize and Rice Protoplasts. Nucl. Acids. Res. 1993, Vol. 21, No. 4, pages 969-975, see entire document.	1,5-7 ----- 1-7, 14-35
Y - A	BAYLEY et al. Exchange of Gene Activity in Transgenic Plants Catalyzed by the Cre-Lox Site-Specific Recombination System. Plant Molecular Biology. 1992, Vol. 18, pages 353-361, see entire document.	1-7, 14-35 ----- 8-13

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

11 JANUARY 2000

Date of mailing of the international search report

03 FEB 2000

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

OUSAMA M-FAIZ ZAGHMOUT

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/19732

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/19732

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I. Claims 1-7, 14-20 are drawn to a composition comprising a population of protoplast library members harboring at least one species of a library of heterologous polynucleotide sequences.

Group II. Claims 8-13 are drawn to a library of protoplast harboring at least one species of a selected shuffled library of heterologous polynucleotide sequences.

Group III. Claims 21-25 are drawn a method for rapid evolution of polynucleotide sequences conferring a desired or predetermined phenotype.

Group IV. Claims 26-34 are drawn to a method for identifying polynucleotide sequences encoding a predetermined phenotype for a plant cell.

Group V. Claim 35 is drawn to a kit for obtaining a polynucleotide encoding a predetermined phenotype.

The inventions listed as groups I-V do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Since the composition comprising a population of protoplast library members harboring at least one species of a library of heterologous polynucleotide sequences, is known in the art as evidenced by the Lyznik et al [Nucleic Acids Research. 1993. Nucleic Acids Research. Vol. 21: 969-975] reference, it does not constitute a special technical feature as defined by PCT Rule 13.2. Groups I-V are directed to composition comprising a population of protoplast library members harboring at least one species of a library of heterologous polynucleotide sequences. They are further directed toward to a library of protoplast harboring at least one species of a selected shuffled library of heterologous polynucleotide sequences, a method for rapid evolution of polynucleotide sequences conferring a desired or predetermined phenotype, a method for identifying polynucleotide sequences encoding a predetermined phenotype for a plant cell, and to a kit for obtaining a polynucleotide encoding a predetermined phenotype. However, since claim 1 lacks novelty, unity of invention is lacking, because composition comprising a population of protoplast library members harboring at least one species of a library of heterologous polynucleotide sequences is known in the art as evidenced by the teaching of Lyznik et al. The cited evidence proves that the technical feature of group I, composition comprising a population of protoplast library members harboring at least one species of a library of heterologous polynucleotide sequences, does not make a contribution over the prior art. The claims are not so linked by a special technical feature within the meaning of the PCT Rule 13.2 so as to form a single inventive concept, accordingly, the unity of invention is lacking among all groups.