(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 August 2020 (13.08.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/163027 A1

(51) International Patent Classification:
A01H 5/00 (2018.01) *C12N 1/00* (2006.01)
A01N 63/00 (2020.01) *C12N 1/20* (2006.01)
A01P 21/00 (2006.01) *C12N 5/04* (2006.01)

(21) International Application Number:
PCT/US2020/012041

(22) International Filing Date:
02 January 2020 (02.01.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/802,805 08 February 2019 (08.02.2019) US
62/802,831 08 February 2019 (08.02.2019) US

(71) Applicant: **NEWLEAF SYMBIOTICS, INC.** [US/US]; 1005 N. Warson Rd., St. Louis, Missouri 63132 (US).

(72) Inventors: **KEROVUO, Janne**; c/o NEWLEAF SYMBIOTICS, INC., 1005 N. Warson Rd., St. Louis, Missouri 63132 (US). **BRYANT, Doug**; c/o NEWLEAF SYMBIOTICS, INC., 1005 N. Warson Rd., St. Louis, Missouri 63132 (US). **VOGAN, Patrick**; c/o NEWLEAF SYMBIOTICS, INC., 1005 N. Warson Rd., St. Louis, Missouri 63132 (US). **BREAKFIELD, Natalie**; c/o NEWLEAF SYMBIOTICS, INC., 1005 N. Warson Rd., St. Louis, Missouri 63132 (US).

(74) Agent: **ROMANO, Charles P.** et al.; Thompson Coburn LLP, One US Bank Plaza, St. Louis, Missouri 63101 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS FOR IDENTIFYING MICROBIAL STRAINS HAVING ENHANCED PLANT COLONIZATION EFFICIENCY

(57) Abstract: Methods of evaluating microorganisms to determine efficiency of colonization of a plant or plant part are described. Methods of identifying genetic elements correlated with colonization efficiency of plant-associated microorganism are also provided. Methods to identify microorganisms for use as inoculants for improved plant yield using the colonization screening methods or the presence of genetic elements associated with colonization efficiency are described. Further methods useful for identification of microorganisms useful as inoculants for improving plant yield are presented.

WO 2020/163027 A1

INTERNATIONAL PATENT APPLICATION FOR METHODS FOR IDENTIFYING MICROBIAL STRAINS HAVING ENHANCED PLANT COLONIZATION EFFICIENCY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This patent application claims the benefit of U.S. 62/802,805, filed February 8, 2019, and U.S. 62/802,831, filed February 8, 2019, which are each incorporated herein by reference in their entireties.

SEQUENCE LISTING STATEMENT

[0002] A sequence listing containing the file named "53907_193367 _ST25.txt" which is 498,169 bytes (measured in MS-Windows®) and created on January 2, 2020, contains 90 protein sequences, and 45 nucleic acid sequences is provided herewith via the USPTO's EFS system, and is incorporated herein by reference in its entirety.

BACKGROUND

[0003] Microbial colonization of plant host cells or tissues is an aspect in both beneficial and detrimental interactions between plant-associated microorganisms and their host. Beneficial interactions include, for example, plant growth promotion by colonizing bacteria or fungi, biopesticidal activity of colonizing microorganisms that protect the host plant from pathogens, and activities of colonizing microorganisms that increase yield of a plant. Detrimental interactions include, for example, colonization of plants by disease causing pathogens. Identification of microorganisms that efficiently colonize plants, plant parts or tissues is useful for development of microbial inoculants for improving plant cultivation. Identification of genes correlated with the ability of a microorganism to colonize a plant host cell or tissue will expand the understanding of plant host/microorganism interactions and allow for manipulation of these interactions to promote interactions that benefit the plant and disrupt interactions that are detrimental to plant growth.

SUMMARY

Methods for identifying one or more genetic elements correlated with colonization efficiency of a plant-associated microorganism that comprise: (i) screening a population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (ii) identifying a first set of strains in said population that

colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment or other strains of said population; (iii) identifying a second set of strains in said population that colonize said plant or plant part at a reduced density as compared to other strains of said population, or at a density that is reduced or not significantly different from that of a non-colonizing control treatment; (iv) comparing the sequences of genetic elements in said first set of strains and said second set of strains; and (v) identifying one or more genetic elements that correlate with colonization efficiency.

Also provided are methods for selecting a microbial strain capable of efficiently colonizing a plant, plant cell or plant part that comprise detecting in the genome of said microbial strain one or more genetic elements that are positively correlated with colonization efficiency.

Further, methods are provided of selecting a microbial strain capable of efficiently colonizing a plant, plant cell, or plant part that comprise detecting the presence of one or more genetic elements in the genome of said microbial strain, wherein said one or more genetic elements (i) comprise a gene selected from the group consisting of bbsG, hdhA, luxQ, bicA, hddC, hddA, fptA, livF, sutR, cdhR, amaB, ssuA, rbn, ftsY, fecA, gpmA_2, ecfG_1, adh, lgt, yfih, cyaA, vgb_3, pimB_2, bmr3_2, and fabD_1, or (ii) encode at least one protein having an amino acid sequence of SEQ ID NO: 1 to SEQ ID NO: 45.

Also provided are methods of identifying a yield-enhancing plant-associated microorganism that comprise: (i) screening a population of plant-associated microorganisms to identify strains having tolerance to desiccation and tolerance to contact with agricultural chemicals, (ii) screening said population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (iii) identifying strains having tolerance to desiccation and contact with agricultural chemicals and strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment; (iii) screening said population of plant-associated microorganisms or said strains identified as having tolerance to desiccation and agricultural chemicals and as having the ability to colonize a plant or plant part to identify strains that efficiently colonize a target plant host; (iv) contacting strains identified as having tolerance to desiccation and agricultural chemicals and that efficiently colonize a plant host with said plant host or a part of said plant host; (v) growing said plant host under field conditions; (vi) determining harvested yield of said plant host; and (vii) identifying strains that improve yield of said plant host.

DETAILED DESCRIPTION

[0004] Methods for identification of plant-associated microorganisms that are capable of efficiently colonizing a plant host are provided and used to identify genetic elements that are correlated with enhanced colonization efficiency.

[0005] The term "and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following embodiments: A, B, and C; A, B, or C; A or C; A or B; B or C; A and C; A and B; B and C; A (alone); B (alone); and C (alone).

[0006] Where a term is provided in the singular, embodiments comprising the plural of that term are also provided.

[0007] As used herein, the terms "include," "includes," and "including" are to be construed as at least having the features or encompassing the items to which they refer while not excluding any additional unspecified features or unspecified items.

[0008] As used herein a "plant-associated microorganism" is a bacterium or fungus that is present in or lives on plants or in soil where plants are grown. Plant-associated microorganisms may be present in the rhizosphere or phyllosphere. Plant-associated microorganisms may also be endophytes that live within a plant or plant part.

[0009] As used herein, the term "desiccation tolerance" is intended to indicate the ability of a plant-associated microorganism to survive under conditions of extreme dryness.

[0010] As used herein, the term "strain" refers to a pure culture of a subject plant-associated microorganism as well as to the progeny or potential progeny of the subject microorganism. The term strain shall include all isolates of such strain.

[0011] As used herein, a "pan-genome" is the entire set of genes for the microbial population being screened in a plant colonization efficiency screen. Thus, a pan-genome may represent the entire set of genes for a particular species, or the entire set of genes in multiple different species of the same genus or even the entire set of genes for multiple species classified in more than a single genus, where the strains in the population are from closely related genera.

[0012] As used herein a "genetic element" refers to an element in a DNA or RNA molecule that comprises a series of adjacent nucleotides at least 20 nucleotides in length and up to 50, 100, 1000, or 10000 or more, nucleic acids in length. A genetic element may

comprise different groups of adjacent nucleic acids, for example, where the genome of a plant-associated microorganism contains introns and exons. The genetic element may be present on a chromosome or on an extrachromosomal element, such as a plasmid. In eukaryotic plant-associated microorganisms, the genetic element may be present in the nucleus or in the mitochondria. In some embodiments, the genetic element is a functional genetic element (*e.g.,* a gene) that encodes a protein.

[0013] As used herein, the terms "homologous"' or "homologue" or "ortholog" refer to related genetic elements or proteins encoded by the genetic elements that are determined based on the degree of sequence identity. These terms describe the relationship between a genetic element or encoded protein found in one isolate, species or strain and the corresponding or equivalent genetic element or protein in another isolate, species or strain. As used herein, a particular genetic element in a first isolate, species or strain is considered equivalent to a genetic element present in a second isolate, species or strain when the proteins encoded by the genetic element in the isolates, species or strains have at least 50 percent identity. Percent identity can be determined using a number of software programs available in the art including BLASTP, ClustalW, ALLALIGN, DNASTAR, SIM, SEQALN, NEEDLE, SSEARCH and the like.

[0014] As used herein, the term “colonize” refers to the ability of a microorganism to grow and reproduce in an environment. A microorganism is considered to colonize a plant or plant part if it can survive and grow on or inside the plant or plant part, including inside a plant cell.

[0015] As used herein, a “population of plant-associated microorganisms” refers to a group of 2 or more strains of genetically related microorganisms. The genetically related microorganisms may, for example, be of the same genus, or of the same species.

[0016] “Colonization efficiency” as used herein refers to the relative ability of a given microbial strain to colonize a plant host cell or tissue as compared to non-colonizing control samples or other microbial strains. Colonization efficiency can be assessed, for example and without limitation, by determining colonization density, reported for example as colony forming units (CFU) per mg of plant tissue, or by quantification of nucleic acids specific for a strain in a colonization screen, for example using qPCR.

[0017] As used herein, a “non-colonizing control treatment” is a treatment that does not contain a strain of a plant-associated microorganism or contains a strain of a plant-associated microorganism that has previously been determined to exhibit low colonization efficiency or no ability to colonize a subject plant or plant part.

[0018] As used herein, a "correlation" is a statistical measure that indicates the extent to which two or more variables, here colonization efficiency and identified genetic elements, occur together. A positive correlation indicates that a microbial strain containing a given genetic element is likely to be an efficient colonizer. A negative correlation indicates that a microbial strain containing a given genetic element is likely to be a poor or inefficient colonizer.

[0019] As used herein, the term "*Methylobacterium*" refers to genera and species in the methylobacteriaceae family, including bacterial species in the *Methylobacterium* genus and proposed *Methylorubrum* genus (Green and Ardley (2018)). *Methylobacterium* includes pink-pigmented facultative methylotrophic bacteria (PPFM) and also encompasses the non-pink-pigmented *Methylobacterium* nodulans, as well as colorless mutants of *Methylobacterium* isolates. For example, and not by way of limitation, "*Methylobacterium*" refers to bacteria of the species listed below as well as any new *Methylobacterium* species that have not yet been reported or described that can be characterized as *Methylobacterium* or *Methylorubrum* based on phylogenetic analysis: *Methylobacterium adhaesivum; Methylobacterium oryzae; Methylobacterium aerolatum; Methylobacterium oxalidis; Methylobacterium aquaticum; Methylobacterium persicinum; Methylobacterium brachiatum; Methylobacterium phyllosphaerae; Methylobacterium brachythecii; Methylobacterium phyllostachyos; Methylobacterium bullatum; Methylobacterium platani; Methylobacterium cerastii; Methylobacterium pseudosasicola; Methylobacterium currus; Methylobacterium radiotolerans; Methylobacterium dankookense; Methylobacterium soli; Methylobacterium frigidaeris; Methylobacterium specialis; Methylobacterium fujisawaense; Methylobacterium tardum; Methylobacterium gnaphalii; Methylobacterium tarhaniae; Methylobacterium goesingense; Methylobacterium thuringiense; Methylobacterium gossipiicola; Methylobacterium trifolii; Methylobacterium gregans; Methylobacterium variabile; Methylobacterium haplocladii; Methylobacterium aminovorans* (*Methylorubrum aminovorans*); *Methylobacterium hispanicum; Methylobacterium extorquens (Methylorubrum extorquens); Methylobacterium indicum; Methylobacterium podarium (Methylorubrum podarium); Methylobacterium iners; Methylobacterium populi (Methylorubrum populi); Methylobacterium isbiliense; Methylobacterium pseudosasae (Methylorubrum pseudosasae); Methylobacterium jeotgali; Methylobacterium rhodesianum (Methylorubrum rhodesianum); Methylobacterium komagatae; Methylobacterium rhodinum (Methylorubrum rhodinum); Methylobacterium longum; Methylobacterium salsuginis (Methylorubrum salsuginis); Methylobacterium marchantiae; Methylobacterium suomiense*

(Methylorubrum suomiense; Methylobacterium mesophilicum; Methylobacterium thiocyanatum (Methylorubrum thiocyanatum); Methylobacterium nodulans; Methylobacterium zatmanii (Methylorubrum zatmanii); Methylobacterium organophilum.

[0020] To the extent to which any of the preceding definitions is inconsistent with definitions provided in any patent or non-patent reference incorporated herein by reference, any patent or non-patent reference cited herein, or in any patent or non-patent reference found elsewhere, it is understood that the preceding definition will be used herein.

[0021] Methods for identification of plant-associated microorganisms that are capable of efficiently colonizing a plant host are provided. In some embodiments, such methods are used to identify genetic elements in said plant-associated microorganisms that are correlated with enhanced colonization efficiency.

[0022] In some embodiments, the plant-associated microorganism is beneficial to a plant host, and the genetic elements correlated with enhanced colonization efficiency can be used to identify other microbial strains that comprise one or more of the genetic elements correlated with enhanced colonization of a plant host. In some embodiments, the plant beneficial microbial strains identified as being able to efficiently colonize a plant host are further screened for additional traits that can contribute the fitness of the plant-associated microbial strain for use as an agricultural inoculant. In some embodiments a population of plant-associated microorganisms is screened, prior to conducting a colonization efficiency screen, for additional traits that can contribute to the fitness of the plant-associated microbial strain for use as an agricultural inoculant. Additional screens that can be used to evaluate the fitness of a particular plant-associated microorganism for use as an agricultural inoculant are also provided and include screens for tolerance to desiccation, tolerance to agricultural chemicals, and screens for growth rate and ease of production when grown in media with varying sources of carbon, nitrogen and other nutrients, such as vitamins or other trace elements.

[0023] In some embodiments, the plant-associated microorganism is detrimental to a plant host, or to a human or animal that consumes the plant or food or feed prepared from said plant that comprises the detrimental plant-associated microorganism. In such embodiments, identified genetic elements correlated with colonization efficiency are used to develop methods to disrupt the function of the detrimental plant-associated microorganisms or prevent such detrimental plant-associated microorganisms from colonizing the plant host. Colonization efficiency screens are provided for determining the ability of a plant-associated microorganism to colonize a host plant or host plant part. In some embodiments, the plant-

associated microorganism is beneficial to a plant, for example as a biostimulant, that improves yield, or a biopesticide that provides protection against plant pests. Microbial biostimulants benefit plants by enhancing nutrition efficiency, abiotic stress tolerance and/or crop quality traits, and include biofertilizers which increase the supply or availability of nutrients for the plant host. Thus, microbial biostimulants include mycorrhizal and non-mycorrhizal fungi, bacterial endosymbionts and rhizobacteria. Non-limiting example of plant-associated bacteria that can be used in the methods provided herein include species of *Actinomycetes, Agrobacterium, Arthrobacter, Alcaligenes, Aureobacterium, Azobacter, Azorhizobium, Azospirillum, Azotobacter, Beijerinckia, Brevibacillus, Burkholderia, Chromobacterium, Clostridium, Clavibacter, Comomonas, Corynebacterium, Curtobacterium, Enterobacter, Flavobacterium, Gluconacetobacter, Gluconobacter, Herbaspirillum, Hydrogenophage, Klebsiella, Luteibacter, Lysinibacillus, Mesorhizobium, Methylobacterium, Microbacterium, Ochrobactrum, Paenibacillus, Pantoea, Pasteuria, Phingobacterium, Photorhabdus, Phyllobacterium, Pseudomonas, Rhizobium, Rhodococcus, Bradyrhizobium, Serratia, Sinorhizobium, Sphingomonas, Streptomyces, Stenotrophomonas, Variovorax, Xanthomonas and Xenorhadbus.* Non-limiting example of plant-associated mycorrhizal and non-mycorrhizal fungi that can be used in the methods provided herein include species of *Acremonium, Alternaria, Ampelomyces, Aspergillus, Aureobasidium, Beauveria, Botryosphaeria, Cladosporium, Cochliobolus, Colletotrichum, Coniothyrium, Embellisia, Epicoccum, Fusarium, Gigaspora, Gliocladium, Glomus, Laccaria, Metarhisium, Muscodor, Nigrospora, Paecilonyces, Paraglomus, Penicillium, Phoma, Pisolithus, Podospora, Rhizopogon, Scleroderma, Trichoderma, Typhula, Ulocladium, and Verticilium.*

[0024] In some embodiments, a plant-associated microorganism in a colonization efficiency screen is detrimental to a plant host, for example as a plant pathogen. Plant pathogenic microorganisms include for example, plant pathogenic fungi, such as species of *Fusarium, Colletotrichum, Septoria, Blumeria, Alternaria, Stagonospora, Stenocarpella, Pythium, Rhizoctonia, Magnaportha, Pyrenophora, Peronospora, Microdochium, Sclerotinia, Sclerospora, Sclerophthora, Phytophthora, Cercospora, Gibberella, Verticillium* and *Pythium*; and plant pathogenic bacteria, including species of *Xanthomonas, Ralstonia, Erwinia, Pseudomonas, Streptomyces, Corynebacterium, Clavibacter, Xylella, Pectobacterium, Dickeya, Pantoea, Serratia,* and *Sphingomonas.* In some embodiments, a plant-associated microorganism is detrimental to humans or other animals that consume a plant part that has been colonized by the microorganism. Plant-associated microorganisms

that are detrimental to humans or other animals include, for example, species of *Salmonella, Listeria, Clostridium*, and *Campylobacter*.

[0025] In some embodiments, a plant-associated microorganism in a plant colonization efficiency screen is epiphytic, living on the surface of a plant or plant part, such as the leaves, roots, flowers, buds, seeds or fruits. In some embodiments, a plant-associated microorganism is an endophyte, living within a plant host for at least a portion of the microorganism's life cycle. In some embodiments, a plant-associated microorganism resides inside a plant host cell for at least a portion of the microorganism's life cycle.

[0026] In some embodiments, a host plant in a plant colonization efficiency screen is a crop plant. Crop host plants include, but are not limited to, corn, soybean, wheat, *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B. juncea*) including Canola varieties, alfalfa, rice, rye, sorghum, millet (e.g., pearl millet (*Pennisetum glaucum*)), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*), sunflower, safflower, tobacco, potato, peanuts, lentils, cotton, sweet potato (*Ipomoea batatus*), cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, papaya, cashew, macadamia, almond, sugar beets, sugarcane, oats, barley, tomatoes, lettuce, green beans, lima beans, peas, cucurbits such as cucumber, cantaloupe, and musk melon, ornamentals, and conifers. Ornamental plant hosts that can be used in a plant colonization efficiency screen include, but are not limited to azalea, hydrangea, hibiscus, roses, tulips, daffodils, petunias, carnation, poinsettia, and chrysanthemum. Conifer host plants that can be used in a plant colonization efficiency screen include, but are not limited to, pines such as loblolly pine, slash pine, ponderosa pine, lodge pole pine, and Monterey pine; Douglas-fir; Western hemlock; Sitka spruce; redwood; true firs such as silver fir and balsam fir; and cedars such as Western red cedar and Alaska yellow-cedar. Turfgrass plant hosts include, but are not limited to, annual bluegrass, annual ryegrass, Canada bluegrass, fescue, bentgrass, wheatgrass, Kentucky bluegrass, orchard grass, ryegrass, redtop, Bermuda grass, St. Augustine grass, and zoysia grass. In some embodiments, the host plant in a plant colonization efficiency screen is a cereal plant selected from the group consisting of a rice, wheat, corn, barley, millet, sorghum, oat, and rye plant or plant part. Host plant parts that can be colonized include, but are not limited to, leaves, stems, flowers, roots, seeds, fruit, tubers, coleoptiles, and the like.

[0027] In some embodiments, an appropriate dose of the plant-associated microorganism for use in a colonization efficiency screen is determined in a dose-response evaluation using one or more strains to compare results using different dose applications. In

some embodiments, an initial dose for evaluation is known from prior use of a strain or a related strain, for example as an agricultural inoculant. In some embodiments the plant-associated microorganism employed in a colonization screen is provided at a lower dose than typically used in agriculture in order to identify strains that would provide an advantage for commercial production. Production costs for agricultural inoculants can be reduced considerably when substantial colonization is achieved using a lower amount of microbial inoculant. In some embodiments, a dose of 10^2 to 10^8 CFU of a microbial strain to be tested for colonization efficiency is applied to a plant or plant part. In some embodiments, a dose of 10^3, 10^4, 10^5, 10^6, or 10^7 CFU is applied to a plant or plant part. In some embodiments, the dose is applied to a plant root, leaf, stem, seed, fruit or flower.

[0028] In some embodiments, a population of plant-associated microorganisms in a colonization screen will comprise a single strain to be assayed and a control strain. In some embodiments, two or more strains will be screened in a colonization assay with or without a control treatment to determine relative colonization efficiency of the strains in the screen. In some embodiments, a population of plant-associated microorganisms in a colonization screen will comprise two or more strains to be tested and a control, where the control can be a treatment where no microorganism is added to a sample, a treatment where a microorganism known or previously determined to be a poor colonizer of the target plant host is used, or a treatment where a microorganism known or previously determined to be an efficient colonizer of the target plant host is used. In some embodiments, both a control lacking added microbial strain and a control microbial strain known to be a poor colonizer may be used. In some embodiments, larger populations of strains are used in a colonization efficiency screen where results of the screen will be used to identify genetic elements associated with colonization efficiency as described herein.

[0029] In some embodiments, a population of plant-associated microorganisms used in a colonization efficiency screen is a population of strains from the same genus and species. In some embodiments, the population contains strains from the same genus, but includes strains of different species within the genus. In some embodiments, the population contains closely related strains from species that have been classified as belonging to different genera. By way of example, it has recently been proposed that some species of bacteria in the genus *Methylobacterium* should be reclassified as belonging to a new genus with the proposed name *Methylorubrum.* The species proposed to belong to the new genus are still highly related to other *Methylobacterium* species and a population of microorganisms for use in a colonization efficiency screen as described herein can contain

species assigned to *Methylobacterium* or *Methylorubrum* or a mixed population with species assigned to one or the other of these general. Similar examples of plant-associated microorganisms that have been assigned to different genera but could be included in a population for screening for colonization efficiency by the methods described herein due their genetic similarity include, for example species of rhizobia, which may be *Rhizobium, Sinorhizobium* or *Mesorhizobium*. Populations of microorganisms for use in the present methods thus can include strains from the same or closely related genera in a single population to be screened for colonization efficiency.

[0030] The population in a plant colonization screen as described herein may be varied depending on factors including, for example, if the results of the screen will be used in association analyses, and the genetic relatedness of strains within the population. In some embodiments, a single strain is a plant-associated microorganism is assayed to determine colonization efficiency. In some embodiments, the colonization efficiency of a strain is determined by comparison to a control sample lacking an added microbial strain. In some embodiments, the colonization efficiency of a strain is determined by comparison to a control sample containing a strain previously identified as either a poor colonizer or an efficient colonizer. In some embodiments, a population will include more than one strain to be screened for colonization efficiency. In some embodiments, a population will include 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or more strains to be screened for colonization efficiency. In some embodiments, the population will include native strains obtained from a plant environment. In some embodiments, the population will include isolates of a strain that has been mutagenized by either directed or random mutagenesis. In some embodiments, a population will include mutagenized isolates of a strain identified as an efficient colonizer and isolates of the strain having even greater colonization efficiency can be identified. In some embodiments, a strain identified as beneficial to a plant host, but lower in colonization efficiency than other genetically related strains is mutagenized to generate a population that can be screened to identify an isolate of the strain that has improved colonization efficiency.

[0031] In some embodiments, results of the colonization assay are used in association analyses and strains in the population are genetically related, such as members of the same microbial species. In some embodiments, results of the colonization assay are used in association analyses and strains in the population are genetically related, such as members of the same microbial species. In some embodiments, strains in the population are genetically related and results of the colonization assay are used in association analyses to identify genetic elements associated with colonization efficiency. In some embodiments, all strains

in the population are members of the same microbial species, as identified for example by comparing the relatedness of genome sequences of the strains. In some embodiments, the strains in the population will have average nucleotide identity (ANI) values of about 95%. In some embodiments, the strains in the population will have ANI values of about 96%. In some embodiments, the strains in the population will have ANI values of about 97 - 99%. In some embodiments, the strains in the population will have ANI values of about 98-99%. In some embodiments, the strains in the population will have ANI values of about 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9%

[0032] In some embodiments, the strains will be members of different species of the same genus, or different species of two or more closely related genera. In some embodiments, strains of different species that find use in the methods herein will have ANI values equal to or greater than 80%, 85%, 90%, and 95%.

[0033] In some embodiments, the population size is sufficient to allow for identification of at least 5 strains that efficiently colonize the plant host or plant host part and at least 5 strains that are poor colonizers of the plant host or plant host part for use in association analyses. In embodiments where the population comprises more distantly related microbial strains, the population will be of a sufficient size to allow for identification of 10 or more strains that efficiently colonize the plant host or plant host part and 10 or more strains that are poor colonizers of the plant host or plant host part. In some embodiments, the population in a plant colonization efficiency screen will comprise 10 or more distinct strains. In some embodiments, the population in a plant colonization efficiency screen will comprise 20 or more distinct strains. In some embodiments, the population in a plant colonization efficiency screen will comprise 50 or more distinct strains. In some embodiments, the population in a plant colonization efficiency screen will comprise 100 or more distinct strains. In some embodiments, the population in a plant colonization efficiency screen will comprise 200 or more distinct strains. In some embodiments, the population in a plant colonization efficiency screen will comprise 300, 400, or 500 or more distinct strains. In some embodiments, each of the strains in the population is screened in a separate experiment. In some embodiments, multiple strains that can be distinguished, for example by the presence of genetic markers or differences in appearance of colonies may be screened in a single experiment.

[0034] In some embodiments, the population of plant-associated microorganisms comprises 100 or more strains of *Methylobacterium* and results of the screen are used in association analyses to identify genetic elements associated with colonization efficiency. In

some embodiments, the plant is a soybean plant. In some embodiments, the plant part treated in a colonization efficiency screen is a soybean seed. In some embodiments, the dose used to treat a plant or plant part in a colonization screen is 10^5 or 10^6 CFU. In some embodiments, the plant part sampled to determine colonization efficiency is a plant shoot. In some embodiments, the plant is corn. In some embodiments, the plant part treated in a colonization efficiency screen is a corn seed. In some embodiments, the dose used in a colonization screen is 10^5 or 10^6 CFU. In some embodiments, the plant part sampled to determine colonization efficiency is a plant root.

[0035] In some embodiments, a population of plant-associated microorganisms comprises two or more strains of *Methylobacterium* and results of colonization efficiency screens are used to identify microbial strains for testing in field studies for use of the strains as agricultural inoculants. In some embodiments, additional screens are employed to further evaluate the fitness of plant-associated microorganisms that efficiently colonize a host plant for use as an agricultural inoculant on the host crop plant and additional crop plants. In some embodiments, the plant is a corn plant. In some embodiments, the plant part treated in a colonization efficiency screen is a corn seed. In some embodiments, the dose used to treat a plant or plant part in a colonization screen is 10^5 or 10^6 CFU. In some embodiments, the plant part sampled to determine colonization efficiency is a plant root. In some embodiments, the plant is a soybean plant. In some embodiments, the plant part treated in a colonization efficiency screen is a soybean seed. In some embodiments, the dose used to treat a plant or plant part in a colonization screen is 10^5 or 10^6 CFU. In some embodiments, the plant part sampled to determine colonization efficiency is a plant shoot.

[0036] The performance of each strain in a population screened for colonization efficiency is evaluated and quantified to provide a dataset that can be used to identify strains having significantly elevated colonization efficiency as compared to a control, and to identify the strains which demonstrate the lowest ability to colonize the plant host.

[0037] In some embodiments, efficiency of colonization is evaluated by inoculating a host plant, plant part or plant tissue with a known quantity of the plant-associated microorganism and assessing the titer of the microorganism on the plant after a period of growth. In other embodiments, a known quantity of the plant-associated microorganism to be tested for colonization efficiency is added to soil in which a host plant is grown, or provided in water or nutrients that are supplied to the host plant. In some embodiments, a plant seed is inoculated, and colonization efficiency is assessed on plant tissue above the soil level, on plant tissue below the soil level, or in plant seed tissue, for example where the

plant-associated microorganism is an endophyte. Titers or genome copies of the plant-associated microorganism are determined and compared to control samples to determine colonization density. In some embodiments, a target host plant is grown in soil that has been treated, for example sterilized, to remove other microorganisms from the sample. In some embodiments, a target plant host is grown in soil which contains a naturally occurring soil microflora. In some embodiments, a target plant host is grown in non-soil media that include calcined clay that has limited background microbiota. In some embodiments, a control sample containing only naturally occurring soil microflora may demonstrate higher colonization densities than some or all of the strains in the population being tested. In some embodiments, the inclusion of a second control treatment is used to ensure that a useful control is available for each treatment. In some embodiments, a second control treatment contains a plant-associated microorganism previously determined to be a poor or inefficient colonizer of the target host plant. In some embodiments, a plant associated-microorganism used in a second control sample is from the same genus or species as the population of plant-associated microorganisms in the screen.

[0038] In some embodiments, colonization efficiency is assessed after the plant has been allowed to grow (*e.g.*, in a controlled environment), for about one to two weeks to allow for germination and seedling growth. In some embodiments, a plant seed is inoculated, and colonization efficiency is assessed by determining the titer of the plant-associated microorganism on plant tissue below the soil level and comparing to a control. In some embodiments, root colonization efficiency is assessed after the plant has been allowed to grow (*e.g.*, in a controlled environment) for at least a week, or alternatively for two to three weeks. In some embodiments, selective media can be used to determine the population of the plant-associated microorganism that has colonized the test plant, plant part or plant tissue. In one embodiment, the plant-associated microorganism is a *Methylobacterium*, and colonizing bacteria can be recovered from the plant sample and plated onto AMS-MC media. After incubation for a sufficient time to allow the *Methylobacterium* to grow, pink colonies indicative of *Methylobacterium* can be counted to determine the number of *Methylobacterium* colonies. In some embodiments, results or scores, for each strain in the population are recorded as CFU per unit of plant tissue, for example CFU per mg of fresh weight.

[0039] In some embodiments, quantification to allow comparison of the colonization efficiency of the population of strains is accomplished using methods other than direct plating to titer the plant-associated microorganisms. In some embodiments, genome

sequencing is used to determine the relative numbers of plant-associated microorganism that have colonized the test plant, plant part or plant tissue. In some embodiments, 16S rRNA gene sequencing for bacteria or 18S rRNA gene sequencing for fungi, is implemented to quantitate and compare colonization efficiency of different strains in the population being screened. In some embodiments, qPCR is used to quantitate the plant-associated microorganisms that have colonized the host plant, plant part or plant tissue using markers that are known or have been specifically developed for the strains in the population being screened. In some embodiments, results are recorded for each strain in the population as copy number where quantification involves sequence analysis or cycle threshold (ct) values where qPCR is employed.

[0040] In some embodiments, strains of plant-associated microorganisms can be labeled with a marker in order to simplify detection and quantification of the microorganism on the host plant or plant part. Useful markers include both selectable and screenable markers. Selectable markers are generally genes which encode proteins that confer resistance to antibiotics, thus allowing detection of strains containing the marker by the ability to grow in media containing antibiotics. Selectable markers useful in the methods described herein include those conferring resistance to kanamycin, gentamycin, ampicillin, and chloramphenicol as well as other antibiotics known to be active against gram negative bacteria. Screenable markers, also referred to as reporter genes, encode proteins that cause a change in visible characteristics of a bacterial colony, for example, a change in color. Examples of screenable markers that find use in the methods described herein are lacZ, GUS, GFP, mcherry, and the like. Markers can be incorporated into the genome of the strains in the population to be screened, or can be provided on a plasmid that is inserted into the plant-associated microorganisms. Insertion of a plasmid can be accomplished for example, by conjugation, electroporation or other transformation methods known in the art. In some embodiments, where markers are employed results are recorded for each strain in the population based on colorimetric, fluorescent, or luminescent activity depending on the marker employed.

[0041] Following quantification of each strain in a population in a colonization efficiency screen, a first set of strains is selected that colonize the plant part or host at enhanced densities as compared to a control. In some embodiments, the strains in the first set will meet a threshold of statistical significance, for example having a p-value of $p < 0.10$, or having a p-value of $p < 0.05$. In some embodiments, a first set of strains will be

determined as those having colonization efficiency scores in the highest 20% of the population, the highest 10% of the population, or the highest 5% of the population.

[0042] Additionally, a second set of strains is identified, where the strains in the second set are inefficient colonizers of the plant host or plant host part used in the colonization efficiency screen. Strains in the second set display quantitatively lower CFU/mg than a control in the colonization efficiency screen. In some embodiments, strains in the second set representing members of the population with the lowest colonization density scores are present at a density at least one log lower than that of the efficient colonizers in the first set of strains. In some embodiments, a second set of strains will be determined as those having colonization efficiency scores in the lowest 20% of the population, the lowest 10% of the population, or the lowest 5% of the population.

[0043] In certain embodiments, genetic elements correlated with colonization efficiency are identified by comparing the sequences of genetic elements in said first set of strains and said second set of strains to identify genetic elements that are correlated with the trait of colonization efficiency. In some embodiments, a genome-wide association study, or whole genome association study is performed to identify the correlated genetic elements.

[0044] In some embodiments where whole genome association is used to identify genetic elements that correlate with colonization efficiency, a pan-genome is generated for the population being tested, for example as described by Page et al. (*Bioinformatics* (2015) 31:3691–3693). In some embodiments, a pan-genome is generated using a higher number of plant-associated microorganisms than is present in the population being tested. In some embodiments, a pan-genome is generated using a population of 50 or more different strains of related plant-associated microorganisms. In some embodiments, a pan-genome is generated using 100 or more different strains of related plant-associated microorganisms. In some embodiments, a pan-genome is generated using sequences of 200 or more different strains of related plant-associated microorganisms. In some embodiments, a pan-genome is generated using sequences of 300, 400, 500 or more, or even greater than 1000 different strains of related plant-associated microorganisms.

[0045] Where the population of plant-associated microorganisms contains strains from multiple different related species or genera, the conditions for generation of the pan genome are modified to account for the lower homology among the genetic elements and/or encoded proteins of the strains in the population. In one embodiment, the population of plant-associated microorganisms contains strains from a single species, and the sequence identity cutoff in a BLASTP analysis is 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identity. In

some embodiments, the population contains strains from multiple different, but related species, and the sequence identity cutoff is reduced depending on the phylogenetic relatedness of the strains in the population. For example, a cutoff of sequence identity of 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% can be used.

[0046] In some embodiments, using the pan-genome as a reference, the presence or absence of each genetic element in the first set of strains (efficient colonizers) and the second set of strains (poor or undetectable colonizers) is determined. The presence and absence scores are used in a correlation analysis to identify the genetic elements that correlate positively with colonization efficiency. In some embodiments, multiple different *p*-values are generated and evaluated in the analysis, for example raw *p*-value, Bonferroni *p*-value, Benjamini-Hochberg *p*-value, and empirical *p*-value. In some embodiments, correlation is established using a statistical significance threshold based on empirical *p*-value where a cutoff of p less than or equal to 0.05 or p less than or equal to 0.10 is used. In some embodiments, scores for sensitivity, where the presence of the gene is used as a determination that a strain is an efficient colonizer, and/or specificity, where the non-presence or absence of the gene is used as an indicator that a strain is a poor colonizer, are also used in the correlation analysis. In some embodiments, a cutoff of greater than or equal to 10, 20 or 25% sensitivity is used. In some embodiments, a cutoff of greater than or equal to 50, 60, 70, or 75% specificity is used. In other embodiments, higher or lower sensitivity and specificity cutoffs are employed.

[0047] In some embodiments, genetic elements identified as correlated with colonization efficiency will have a predicted function based on homology to previously identified gene or protein sequences. In some embodiments, genetic elements identified as correlated with colonization efficiency will encode a protein annotated as a "hypothetical protein", either from literature-based homologs annotated as such, or based on lack of homology to any publicly known sequence.

[0048] In certain embodiments, genetic elements positively correlated with plant colonization efficiency are used to identify additional plant-associated microorganisms that contain the genetic elements, for example as a prescreen for selection of strains to be tested for use as an agricultural inoculant. In this manner, strains are identified that, based on the presence of the correlated genes, are capable of growth and reproduction on a target host plant or host plant part. In some embodiments, the additional plant-associated microorganisms will be identified by the presence of two or more genetic elements that are positively correlated with colonization efficiency.

[0049] Detection of genetic elements positively correlated with plant colonization efficiency can be accomplished by a nucleic acid detection technique. In certain embodiments, the nucleic acid detection technique detects a sequence encoding a protein of SEQ ID NO:1 to SEQ ID NO: 45, a sequence encoding a protein having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 91 to SEQ ID NO: 135, or a nucleic acid sequence having at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 46 to SEQ ID NO: 90 or a fragment thereof. Applicable nucleic acid detection techniques include nucleic acid hybridization with one or more of SEQ ID NO: 46 to SEQ ID NO: 90 provided herein of fragments thereof (*e.g.*, fragments comprising at least 15, 18, 20, 50, or 100 to about 500 or more nucleotides of SEQ ID NO: 46 to SEQ ID NO: 90). Appropriate stringency conditions which promote DNA hybridization are, for example, 6X sodium chloride/sodium citrate (SSC) at about 45^{0}C, followed by a wash of 2X SSC at 50^{0}C. Such conditions are known to those skilled in the art and can be found, for example in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989). Salt concentration and temperature in the wash step can be adjusted to alter hybridization stringency. For example, conditions may vary from low stringency of about 2X SSC SSC at 40^{0}C to moderately stringent conditions of about 2X SSC at 50^{0}C to high stringency conditions of about 0.2X SSC at 50^{0}C. Other nucleic acid detection techniques that can be used to detect genetic elements positively correlated with plant colonization efficiency include PCR amplification using primers designed from protein and/or nucleic acid sequences provided herein, or direct genome sequencing. Other nucleic acid detection techniques that find use in detection of sequences in a target microorganism include polymerase chain reaction, branched DNA, ligase chain reaction, transcription mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), nanopore-, mass spectroscopy, or direct sequencing based methods, or any combination thereof. Analysis of genome sequences to detect genetic elements positively correlated with plant colonization efficiency can be by comparison of nucleic acid sequences to sequences of genetic elements positively correlated with plant colonization efficiency provided herein, or by analysis of proteins encoded by nucleic acid sequences present in the genomes of plant-associated microorganisms. Sequences of genetic elements or proteins identified in this manner can be compared using standard nucleic acid protein sequence and analysis tools, including for example, BLAST, pFAM, ClustalW, ALLALIGN, DNASTAR, SIM, SEQALN, NEEDLE, SSEARCH, and the like.

[0050] In some embodiments, presence of a genetic element associated with plant colonization is detected where a genetic element in a plant-associated microorganism of interest encodes a protein having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity or more to a protein encoded by a genetic element correlated with colonization efficiency. In certain embodiments, the genetic element comprises a gene that encodes a protein having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% sequence identity or more to a protein encoded by SEQ ID NO: 1 to SEQ ID NO: 45. In some cases where the plant-associated microorganism of interest is more distantly related to strains in the population screened for colonization efficiency, identity to a genetic element correlated with colonization efficiency may be less than 50%, for example, 40% or even 30%. In certain embodiments, the genetic element comprises a gene that encodes a protein having 30% to 50% sequence identity to a protein encoded by SEQ ID NO: 1 to SEQ ID NO: 45. In certain embodiments, a genetic element correlated with colonization efficiency can encode a protein with the biochemical activity (*e.g.,* transcription regulatory, enzymatic, transport, receptor, and/or binding activity) of a gene as set forth and annotated in Table 2. In certain embodiments, a genetic element correlated with colonization efficiency can encode a protein with the biochemical activity (*e.g.,* transcription regulatory, enzymatic, transport, receptor, and/or binding activity) of a gene encoding a protein of SEQ ID NO: 1 to SEQ ID NO: 45. In certain embodiments, a genetic element correlated with colonization efficiency can encode a protein with any of the aforementioned sequence identities to SEQ ID NO: 1 to SEQ ID NO: 45 and the biochemical activity (*e.g.,* transcription regulatory, enzymatic, transport, receptor, and/or binding activity) of a gene encoding a protein of SEQ ID NO: 1 to SEQ ID NO: 45.

[0051] In some embodiments, presence of a genetic element associated with plant colonization is detected where a genetic element in a plant-associated microorganism of interest comprises a nucleic acid sequence having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NO: 46 to SEQ ID NO:90. In some cases where the plant-associated microorganism of interest is more distantly related to strains in the population screened for colonization efficiency or to the *Methylobacterium* strains which contain any one of SEQ ID NO: 46 to SEQ ID NO:90, sequence identity to a genetic element correlated with colonization efficiency may be less than 50%, for example, 40% or even 30%. In certain embodiments, the genetic element comprises a gene that comprises a nucleic acid sequence having 30% to 50% sequence identity to any one of SEQ ID NO: 46 to SEQ ID NO:90.

[0052] In some embodiments, presence of a genetic element associated with plant colonization is detected where a genetic element in a plant-associated microorganism of interest encodes a protein having at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NO: 91 to SEQ ID NO: 135. In some cases where the plant-associated microorganism of interest is more distantly related to strains in the population screened for colonization efficiency or to the *Methylobacterium* strains which contain genetic elements that encode any one of SEQ ID NO: 91 to SEQ ID NO: 135, sequence identity to a protein encoded by a genetic element correlated with colonization efficiency may less than 50%, for example, 40% or even 30% to any one of SEQ ID NO: 91 to SEQ ID NO: 135. In certain embodiments, the genetic element comprises a gene that encodes a protein having 30% to 50% sequence identity to a protein encoded by any one of SEQ ID NO: 46 to SEQ ID NO: 90 (*i.e.,* the protein of SEQ ID NO: 91 to SEQ ID NO: 135, respectively). In certain embodiments, a genetic element correlated with colonization efficiency can encode a protein with any of the aforementioned sequence identities to SEQ ID NO: 91 to SEQ ID NO: 135 and the biochemical activity (*e.g.,* transcription regulatory, enzymatic, transport, receptor, and/or binding activity) of a protein of SEQ ID NO: 91 to SEQ ID NO: 135.In some embodiments, presence of a genetic element associated with plant colonization is detected where a genetic element in a plant-associated microorganism is identified, for example by sequence comparison to known DNA or protein sequence database, as encoding a protein selected from the proteins listed in Table 1 below. Useful methods to compare sequences include, for example, TBLASTN, where the protein sequences are aligned to a nucleotide database in all six reading frames, and BLASTP, where a protein query is used to search a protein database directly.

Table 1

Gene name	**Annotation**
cdhR	HTH-type transcriptional regulator CdhR
amaB	N-carbamoyl-L-amino acid hydrolase
rbn	Ribonuclease BN
bbsG	(R)-benzylsuccinyl-CoA dehydrogenase
fecA	Vitamin B12 transporter BtuB
luxQ	Autoinducer 2 sensor kinase/phosphatase LuxQ
mdlC	Benzoylformate decarboxylase
hddA	D-glycero-alpha-D-manno-heptose 7-phosphate kinase

fptA	Fe(3+)-pyochelin receptor
livF	High-affinity branched-chain amino acid transport ATP-binding protein LivF
hddC	D-glycero-alpha-D-manno-heptose 1-phosphate guanylyltransferase
sutR	HTH-type transcriptional regulator SutR
ftsY	Signal recognition particle receptor FtsY
bicA	C4-dicarboxylic acid transporter DauA
hdhA	7-alpha-hydroxysteroid dehydrogenase
ssuA	Putative aliphatic sulfonates-binding protein
uvrA	UvrABC system protein A
gpmA_2	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase
ecfG_1	ECF RNA polymerase sigma factor EcfG
adh	putative zinc-binding alcohol dehydrogenase
lgt	Prolipoprotein diacylglyceryl transferase
yfih	Laccase domain protein
cyaA	Adenylate cyclase 1
vgb_3	Virginiamycin B lyase
pimB_2	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase
bmr3_2	Multidrug resistance protein 3
fabD_1	Malonyl CoA-acyl carrier protein transacylase

[0053] In some embodiments, protein sequences identified herein are used to identify genetic elements associated with plant colonization in a microbial strain of interest. Provided herein as SEQ ID NO: 1 to SEQ ID NO: 45 are consensus sequences representing *Methylobacterium* proteins encoded by genetic elements associated with colonization of soybean phyllosphere or corn rhizosphere. In some embodiments, SEQ ID NO: 1 to SEQ ID NO: 45 are used to identify additional *Methylobacterium* strains having genetic elements associated with plant colonization efficiency.

[0054] In some embodiments, additional plant-associated microorganisms will be identified by the presence of 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more genetic elements that are positively correlated with colonization efficiency. In some embodiments, a plant-associated microorganism for use on a particular crop host plant or plant part will be selected based on the presence of the genetic elements that are most closely correlated with colonization efficiency of the target plant host. In some embodiments additional plant-associated microorganisms will be selected for use in agricultural testing that have the highest number of correlated genes identified.

[0055] In some embodiments additional plant-associated microorganisms that are selected based on the presence of genetic elements that correlate with colonization efficiency will be from a species that was represented in a population used in the colonization efficiency screen. In some embodiments, additional plant-associated microorganisms that are selected are from more distantly related species. In some embodiments where the microorganisms are more distantly related, the presence of a higher number of genetic elements associated with colonization efficiency may be required to identify additional plant-associated microorganisms capable of efficient plant colonization.

[0056] In some embodiments, a plant-associated microorganism having enhanced colonization efficiency will be obtained by genetically transforming a target strain to contain genetic elements that correlate with colonization efficiency. In some embodiments, transformation of *Methylobacterium* strains can be accomplished by electroporation or conjugation to transfer of vectors, gene-containing fragments, or expression constructs from one *Methylobacterium* strain to a second bacterial strain (*e.g.,* a second *Methylobacterium* strain) as described, for example in co-pending application 62/760,092. In some embodiments, a plant-associated microorganism having enhanced colonization efficiency is genetically transformed to include genetic elements that can confer pest tolerance (*e.g.,* insect, nematode, and/or fungal pathogen tolerance) or herbicide tolerance. In some embodiments, transformation of an enhanced colonization efficiency *Methylobacterium* strain to include genetic elements that can confer pest tolerance (*e.g.,* insect, nematode, and/or fungal pathogen tolerance) or herbicide tolerance as described, for example in co-pending application 62/694,775.

[0057] In some embodiments where the plant-associated microorganism identified as capable of efficiently colonizing a plant is beneficial to a plant host, additional screens are employed to further evaluate the fitness of the plant-associated microorganism for use as an agricultural inoculant. In some embodiments, such further evaluations are conducted prior to or subsequent to screening a population of plant-associated microorganisms for colonization efficiency.

[0058] In some embodiments a screen for desiccation tolerance is employed. Methods for identifying desiccation tolerant microorganisms include screening a population of microorganisms for viability after a period of drying, for example, in one embodiment drying under a laminar flow hood, and comparing viability to other tested strains. In some embodiments, plant-associated microorganisms can be dried directly from the growth medium, for example, in one embodiment, dried in petri dishes or microtiter plates. In some

embodiments, the microorganisms are grown in media having a single carbon source, dried in the minimal media and rehydrated in a rich nutrient media. In some embodiments, plant-associated microorganisms are coated on seeds and allowed to dry and tested for viability after a period of storage on dry seeds. In some embodiments, microorganisms are stored on dry seeds for anywhere from one day to three weeks, including 2 days, 5 days, 1 week, 2 weeks, and 3 weeks or more before testing for viability. In some embodiments, microorganisms are stored on dry seeds for greater than 4 weeks prior t testing for viability. In some embodiments, plant-associated microorganisms are tested for production of exopolysaccharide (EPS) which has been shown to be involved in protection from desiccation (Gasser et al. 2009, FEMS Microbiol Ecol 70:142-150).

[0059] In some embodiments, a screen for the ability of a plant-associated microorganism to tolerate the presence of commonly used agricultural chemicals is used. In some embodiments, microorganisms to be tested for tolerance to agricultural chemicals will be grown in liquid media and spotted onto solid media plates containing the agricultural chemicals. In some embodiments, the agricultural chemicals in such a screen will include herbicides, for example one or more of the following acetochlor, clethodim, dicamba, flumioxazin, fomesafen, glyphosate, glufosinate, mesotrione, quizalofop, saflufenacil, sulcotrione, and 2,4-D. In some embodiments, the agricultural chemicals in such a screen will include fungicides, for example one or more of the following acibenzolar-S-methyl, azoxystrobin, benalaxyl, bixafen, boscalid, carbendazim, cyproconazole, dimethomorph, epoxiconazole, fluopyram, fluoxastrobin, flutianil, flutolanil, fluxapyroxad, fosetyl-Al, ipconazole, isopyrazam, kresoxim-methyl, mefenoxam, metalaxyl, metconazole, myclobutanil, orysastrobin, penflufen, penthiopyrad, picoxystrobin, propiconazole, prothioconazole, pyraclostrobin, sedaxane, silthiofam, tebuconazole, thifluzamide, thiophanate, tolclofos-methyl, trifloxystrobin, and triticonazole. In some embodiments, the agricultural chemicals in such a screen will include insecticides and/or nematicides, including, for example abamectin, aldicarb, aldoxycarb, bifenthrin, carbofuran, chlorantraniliporle, chlothianidin, cyfluthrin, cyhalothrin, cypermethrin, deltamethrin, dinotefuran, emamectin, ethiprole, fenamiphos, fipronil, flubendiamide, fosthiazate, imidacloprid, ivermectin, lambda-cyhalothrin, milbemectin, nitenpyram, oxamyl, permethrin, tioxazafen, spinetoram, spinosad, spirodichlofen, spirotetramat, tefluthrin, thiacloprid, thiamethoxam, tioxazafen, and thiodicarb. In some embodiments, the agricultural chemicals in such a screen will include biocides, such as isothiazolinones, including for example 1,2 Benzothiazolin-3-one (BIT), 5-Chloro-2-methyl-4-isothiazolin-3-

one (CIT), 2-Methyl-4-isothiazolin-3-one (MIT), octylisothiazolinone (OIT), dichlorooctylisothiazolinone (DCOIT), and butylbenzisothiazolinone (BBIT); 2-Bromo-2-nitro-propane-1,3-diol (Bronopol), 5-bromo-5-nitro-1,3-dioxane (Bronidox), Tris(hydroxymethyl)nitromethane, 2,2-Dibromo-3-nitrilopropionamide (DBNPA), and alkyl dimethyl benzyl ammonium chlorides. In some embodiments, the agricultural chemicals in such a screen will include any combination of fungicides, herbicides, insecticides nematicides, and biocides.

[0060] In some embodiments, a screen for the ability of a plant-associated microorganism to grow robustly and to a high titer in media comprising varying sources, concentrations and combinations of carbon, nitrogen and other nutrients, including one or more of the following vitamins or other trace elements, is employed. Such screens find particular interest, for example, where a desirable plant-associated microorganism is known to have a relatively slow growth rate.

[0061] Microbial strains identified as capable of efficiently colonizing a plant host in a screen as defined herein and/or identified as containing genes associated with plant colonization using the methods provided herein, are tested under agricultural field conditions to identify microbial strains that confer increased yield to inoculated host plants. In some embodiments, microbial strains capable of efficiently colonizing a plant host and/or containing genes associated with enhanced colonization efficiency are applied using a lower inoculum dose than typically used in agricultural applications for related strains not selected as having enhanced colonization efficiency. In some embodiments, a dose for a microbial strain capable of efficiently colonizing a plant host will be 95% lower than a dose used for a related inoculant. In some embodiments, a dose for a microbial strain capable of efficiently colonizing a plant host will be 90% lower than a dose used for a related inoculant. In some embodiments, a dose for a microbial strain capable of efficiently colonizing a plant host will be 85%, 80% or 75% lower than a dose used for a related inoculant. In some embodiments, a dose for a microbial strain capable of efficiently colonizing a plant host will be anywhere from about 25% to 75% lower than a dose used for a related inoculant.

[0062] In some embodiments, agricultural field tests are conducted in a microplot to evaluate yield enhancement of a large number of strains. In some embodiments, the size of a microplot is two 30-inch rows, each 10 feet in length, although many variations are possible. In some embodiments, a microplot trial will be conducted at 2 or more sites with four or more replications at each site. In some embodiments, a microplot trial will be conducted at 4 or more sites with four, five or six or more replications at each site. In some embodiments,

the sites selected for the microplot trial will be geographically diverse to allow for yield analysis under different environmental conditions. In some embodiments, 40 or more strains are evaluated in a microplot to identify yield-enhancing strains. In some embodiments, 50, 60, 70, 80, 90, 100 or more strains are tested for yield enhancing properties in a microplot trial.

[0063] In some embodiments, large-scale conventional field trials are conducted to evaluate and/or confirm yield-enhancing capabilities of strains having enhanced colonization efficiency and/or identified as containing genes associated with plant colonization. In some embodiments, the size of a conventional field trial is four 30-inch rows, each 20 feet or 40 feet in length, although many variations are possible. In some embodiments, a large-scale field trial will be conducted at four or more sites with six or more replications at each site. In some embodiments, the sites selected for the conventional field trials will be geographically diverse to allow for yield analysis under different environmental conditions. In some embodiments, fewer strains are evaluated in a large-scale field trial as compared to the number of strains that are evaluated in a microplot. In some embodiments, 2 or more, 5 or more, 10 or more strains, 15 or more, or 20 or more strains are evaluated in a large-scale field trial.

[0064] In some embodiments, candidate yield-enhancing microbial strains will have additional traits that make them amenable to use in agriculture, including for example desiccation tolerance, tolerance to one or more agricultural chemicals, and robust growth. In some embodiments, the yield-enhancing candidate microbial strains are identified as a hit in one or more screens for desiccation tolerance, agricultural chemical tolerance or robust growth as described herein. In some embodiments, the yield-enhancing candidate microbial strains are identified as hits in desiccation tolerance and agricultural chemical tolerance screens. In some embodiments, the yield-enhancing candidate microbial strains are identified as tolerant to one or more fungicides, herbicides, insecticides nematicides, and biocides.

[0065] In some embodiments, the plant associated microbial strain is a *Methylobacterium* strain. In some embodiments, a *Methylobacterium strain* evaluated in plant field trials will demonstrate a high degree of fitness for use as agricultural inoculants as the result of enhanced colonization efficiency and tolerance to conditions and chemicals that are common to various plant inoculation practices, including for example seed treatment. The *Methylobacterium* strains identified as providing for increased yield in treated plants and having enhanced colonization efficiency when the seed or soil near the seed is treated

with such strains, will have higher population densities on target plants or plant parts, reduced costs of application due to the ability to use less inoculum, or enhanced tolerance to agricultural environments.

[0066] Compositions useful for treatment of plants or soil in which plants are grown are also provided. In some embodiments, *Methylobacterium* strains identified as providing for increased yield in treated plants and having enhanced colonization efficiency will find use in agriculture as inoculants for treatments of plants or plant parts. In some embodiments, an effective amount of the strain having enhanced colonization efficiency used in treatment of seeds or plant parts is provided in a composition having a *Methylobacterium* titer of at least about $1x10^6$ colony-forming units per milliliter, at least about $5x10^6$ colony-forming units per milliliter, at least about $1x10^7$ colony-forming units per milliliter, at least about 5×10^8 colony-forming units per milliliter, at least about 1×10^9 colony-forming units per milliliter, at least about 1×10^{10} colony-forming units per milliliter, or at least about 3×10^{10} colony-forming units per milliliter. In some embodiments, an effective amount of the *Methylobacterium* strain having enhanced colonization efficiency used in treatment of seeds or plant parts is provided in a composition with the *Methylobacterium* at a titer of about least about $1x10^6$ colony-forming units per milliliter, at least about $5x10^6$ colony-forming units per milliliter, at least about $1x10^7$ colony-forming units per milliliter, or at least about 5×10^8 colony-forming units per milliliter to at least about 6×10^{10} colony-forming units per milliliter of a liquid or an emulsion. In some embodiments, an effective amount of the *Methylobacterium* strain having enhanced colonization efficiency used in treatment of seeds or plant parts is provided in a composition with the *Methylobacterium* titer of at least about $1x10^6$ colony-forming units per gram, at least about $5x10^6$ colony-forming units per gram, at least about $1x10^7$ colony-forming units per gram, or at least about 5×10^8 colony-forming units per gram to at least about 6×10^{10} colony-forming units of *Methylobacterium* per gram of the composition. In some embodiments, an effective amount of a *Methylobacterium* strain having enhanced colonization efficiency is provided in a composition with a *Methylobacterium* titer of at least about $1x10^6$ colony-forming units per gram, at least about $5x10^6$ colony-forming units per gram, at least about $1x10^7$ colony-forming units per gram, or at least about $5x10^8$ colony-forming units per gram to at least about $6x10^{10}$ colony-forming units of *Methylobacterium* per gram of particles in the composition containing the particles, wherein the particles comprise a solid substance wherein a mono-culture or co-culture of a *Methylobacterium* strain having enhanced colonization efficiency is adhered thereto. In some embodiments, an

effective amount of a *Methylobacterium* strain provided herein can be a composition with a *Methylobacterium* titer of at least about $1x10^6$ colony-forming units per mL, at least about $5x10^6$ colony-forming units per mL, at least about $1x10^7$ colony-forming units per mL, or at least about 5×10^8 colony-forming units per mL to at least about 6×10^{10} colony-forming units of *Methylobacterium* per mL in a composition comprising an emulsion wherein a mono-culture or co-culture of a *Methylobacterium* strain having enhanced colonization efficiency is adhered to a solid. In some embodiments, an effective amount of a *Methylobacterium* strain having enhanced colonization efficiency can be provided in a composition with a *Methylobacterium* titer of at least about $1x10^6$ colony-forming units per mL, at least about $5x10^6$ colony-forming units per mL, at least about $1x10^7$ colony-forming units per mL, or at least about 5×10^8 colony-forming units per mL to at least about 6×10^{10} colony-forming units of *Methylobacterium* per mL of in a composition comprising an emulsion wherein a mono-culture or co-culture of a *Methylobacterium* strain having enhanced colonization efficiency is provided therein or grown therein.

[0067] An effective amount of a *Methylobacterium* strain having enhanced colonization efficiency provided in a treatment of a seed or plant part is an amount that results in an increase in the colonization density of the *Methylobacterium* strain on a plant grown from the treated seed or plant comprising the treated plant part, and a resulting increase in plant performance, for example as measured by plant yield. In some embodiments, an effective amount of a *Methylobacterium* strain having enhanced colonization efficiency provided in a treatment of a seed or plant part will be lower than the amount used for a similar inoculant that is not an efficient colonizer, and is at least about 10^3, 10^4, 10^5, 10^6, 10^7, or 10^8 CFU per seed or treated plant part. In some embodiments, the effective amount of *Methylobacterium* provided in a treatment of a seed or plant part is an amount where the CFU per seed or treated plant part will exceed the number of CFU of any resident naturally occurring *Methylobacterium* strain by at least 5-, 10-, 100-, or 1000-fold. In some embodiments, the effective amount of *Methylobacterium* provided in a treatment of a seed or plant part is an amount where the CFU per seed or treated plant part will exceed the number of CFU of any resident naturally occurring *Methylobacterium* by at least 2-, 3-, 5-, 8-, 10-, 20-, 50-, 100-, or 1000-fold.

EXAMPLES

[0068] The following examples are included to demonstrate certain embodiments of the disclosure. However, those of skill in the art should, in light of the instant disclosure,

appreciate that many changes can be made in the specific embodiments that are disclosed, while still obtaining like or similar results, without departing from the scope of the disclosure.

Example 1 Dose Response Evaluation

[0069] To determine an appropriate target concentration of *Methylobacterium* for use in a colonization efficiency screen, seven strains previously identified as either having the ability to colonize soybean significantly at 10^6 CFU/seed (4 strains) or as demonstrating poor colonization of soybean seeds when inoculated at 10^6 CFU/seed (3 strains), were evaluated. Soybean seeds were treated at both 10^5 and 10^6 CFU/seed with each of the strains. Three repetitions were planted with 6 seeds per treatment per repetition.

[0070] Results demonstrate that inoculation at a target seed titer of 10^5 CFU/seed can be used for identification of *Methylobacterium* strains with ability to colonize soybean shoots at a significantly higher density than control treatments and other strains previously shown to be poor colonizers of soybean shoots.

Example 2 Soybean Phyllosphere Colonization Efficiency Screen

[0071] *Methylobacterium* strains that colonize the shoot surfaces of soy most densely when applied to seed at a dose of 10^5 CFU/seed (a reduction from the 10^6 CFU/seed levels used previously in field trials) were identified as follows.

[0072] *Methylobacterium* strains were tested for their ability to efficiently colonize soybean shoots. Soybean seeds were treated with *Methylobacterium* strains at a target seed titer of 10^5 CFU/seed. Flo-Rite 1706 polymer was used to stick microbe to seed. Each experiment included 10 *Methylobacterium* strains, plus an untreated control treatment (UTC) and a strain that was shown in the past to have limited ability to colonize soybean phyllosphere (NLS0400, the "negative control"). In each experiment, two seeds per pot were planted in unamended field soil, with 20 pots per treatment level in a randomized complete block design, resulting in a total of 240 pots per experimental run. Plants were grown for 2 weeks in a greenhouse at 25^0C with regular watering and no fertilizer. At harvest, the two plants from each pot were cut at ~1 cm above the soil surface and placed into a 50 mL conical tube with 15 mL of 0.9% saline solution. Ten pots per treatment were sampled. Each tube was weighed before and after plant sampling to quantify plant fresh weight. Samples were vortexed for 15 minutes, then placed into an ultrasonic bath for 10 minutes. Samples were then plated onto AMS-MC using an easySpiral automatic diluter and plater (Interscience, Inc.) at 5 dilutions, and plates were incubated for 8 days at 30°C. Plates were

counted using a Scan 4000 automatic colony counter (Interscience, Inc.) to quantify the number of pink colonies. Results were recorded as the number of CFUs per mg of plant fresh weight.

[0073] Colonization density data (CFU/mg) were used to compare the strains in each run to the untreated control. A Mann-Whitney *U*-test was used to generate *p*-values comparing each treatment to the untreated control. The threshold for statistical significance used in this screen was $p < 0.05$. Strains with significantly greater CFU/mg than the UTC were classified as "hits" based on which treatments were significantly higher than the negative control at $p < 0.05$ using a Mann-Whitney *U*-test.

[0074] In some runs, the untreated control showed an unusually high value. NLS0400 was used as the standard for statistical comparison in any run in which 2 treatments or more showed mean CFU/mg lower than the UTC. Typically, hits colonized the shoot surfaces of soy at a rate that was 0.7-1.3 logs more CFUs per mg of plant fresh weight than the UTC or poorly-colonizing strains. Strains were considered poor colonizers and called "non-hits" if they displayed quantitatively lower colonization than the negative or untreated control. In one test, 380 individual strains were screened. Ninety strains were classified as hits, and sixty-five strains were classified as poor colonizers or "non-hits".

Example 3 Association Analysis

[0075] Genomes of the strains screened in the colonization efficiency were assembled and putative genes identified, then assigned a putative function by sequence analysis to databases of known genes and gene signatures. A pan-genome for *Methylobacterium* was constructed as described by Page *et al.* (Roary: rapid large-scale prokaryote pan genome analysis, *Bioinformatics* (2015) 31:3691–3693) except that genome sequences from greater than 1000 different species of *Methylobacterium* were assembled and used to construct the pan-genome as opposed to the single *Salmonella* species described by Page et al.

[0076] Strains that were identified as the most efficient colonizers or "hits" (as defined by *p*-value in the screen described above) and the poorest colonizers, "non-hits", were compared to determine the presence or absence in each strain of each genetic element in the pan-genome. For this analysis, translated genes were clustered across strains using BLASTP with a sequence identity of at least 50% to identify homologous genetic elements across genomes. These results are used to determine which genetic elements are the same or different across strains, leading to a score for each genetic element as present or absent in a given strain. The presence/absence scores were used in a correlation analysis to identify genetic elements that correlate positively with colonization efficiency as described by

Brynildsrud *et al.* (Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary, *Genome Biology* (2016) 17:238).

[0077] The steps in the process are as follows. Correlated genetic elements were collapsed so that genes that are typically inherited together, for example genes on the same plasmid, were combined into a single unit. Each genetic element in the pan-genome received a null hypothesis of no association to the trait. A Fisher's exact test was performed on each genetic element with the assumption that all strains had a random and independently distributed probability for exhibiting each state, i.e. presence or absence of the genetic element. To control spurious associations due to population structure, the pairwise comparisons algorithm was applied using a phylogenetic tree of the *Methylobacterium* genus, constructed using the same genome sequences described above. Empirical *p*-value was computed using label-switching permutations, i.e. the test statistic was generated over random permutations of the phenotype data. The genetic elements that were significantly positively correlated with colonization efficiency were identified based on *p* value using a threshold for statistical significance of *p* less than or equal to 0.05. Sensitivity and specificity cutoffs were also employed. For sensitivity, i.e. using the presence of the gene as a determination of a strain as an efficient colonizer, a cutoff of greater than or equal to 25% was used. For specificity, i.e. using the non-presence of the gene as an indicator of a strain as a poor colonizer, a cutoff of greater than or equal to 75% was used.

[0078] Gene elements that are positively correlated with *Methylobacterium* colonization of soybean phyllosphere are shown in Table 2 below.

Table 2

Gene name	**Consensus Protein SEQ ID NO:**	**Representative gene sequences Nucleotide/Protein**	**Annotation**	**Sensitivity**	**Specificity**	**p-value**
cdhR_1	1	SEQ ID NO:46/SEQ ID NO: 91	hypothetical protein	39.56	90.91	0.003
group_37721	2	SEQ ID NO: 47/SEQ ID NO: 92	hypothetical protein	37.36	81.82	0.005
amaB_5	4	SEQ ID NO: 48/SEQ ID NO: 93	N-carbamoyl-L-amino acid hydrolase	27.47	86.36	0.005
group_53552	4	SEQ ID NO: 49/SEQ ID NO: 94	hypothetical protein	26.37	89.39	0.008
rbn	5	SEQ ID NO: 50/SEQ ID NO: 95	Ribonuclease BN	26.37	93.94	0.016
group_6	6	SEQ ID NO:	hypothetical protein	38.46	78.79	0.018

Gene name	Consensus Protein SEQ ID NO:	Representative gene sequences Nucleotide/Protein	Annotation	Sensitivity	Specificity	p-value
798		51/SEQ ID NO: 96				
group_57813	7	SEQ ID NO: 52/SEQ ID NO: 97	hypothetical protein	27.47	92.42	0.019
bbsG_1	8	SEQ ID NO: 53/SEQ ID NO: 98	(R)-benzylsuccinyl-CoA dehydrogenase	41.76	84.85	0.022
fecA	9	SEQ ID NO: 54/SEQ ID NO: 99	Vitamin B12 transporter BtuB	28.57	86.36	0.022
luxQ_3	10	SEQ ID NO: 55/SEQ ID NO: 100	Autoinducer 2 sensor kinase/phosphatase LuxQ	30.77	84.85	0.028
group_7164	11	SEQ ID NO: 56/SEQ ID NO: 101	hypothetical protein	38.46	86.36	0.028
mdlC	12	SEQ ID NO: 57/SEQ ID NO: 102	Benzoylformate decarboxylase	45.05	77.27	0.029
hddA	13	SEQ ID NO: 58/SEQ ID NO: 103	D-glycero-alpha-D-manno-heptose 7-phosphate kinase	34.07	83.33	0.030
fptA	14	SEQ ID NO: 59/SEQ ID NO: 104	Fe(3+)-pyochelin receptor	38.46	77.27	0.031
livF_5	15	SEQ ID NO: 60/SEQ ID NO: 105	High-affinity branched-chain amino acid transport ATP-binding protein LivF	29.67	84.85	0.031
group_45430	16	SEQ ID NO: 61/SEQ ID NO: 106	hypothetical protein	28.57	92.42	0.031
group_29067	17	SEQ ID NO: 62/SEQ ID NO: 107	hypothetical protein	26.37	92.42	0.031
hddC	18	SEQ ID NO: 63/SEQ ID NO: 108	D-glycero-alpha-D-manno-heptose 1-phosphate guanylyltransferase	30.77	84.85	0.032
sutR_2	19	SEQ ID NO: 64/SEQ ID NO: 109	HTH-type transcriptional regulator SutR	37.36	78.79	0.033
ftsY	20	SEQ ID NO: 65/SEQ ID NO: 110	Signal recognition particle receptor FtsY	40.66	78.79	0.033
bicA	21	SEQ ID NO: 66/SEQ ID NO:	C4-dicarboxylic acid transporter DauA	42.86	77.27	0.034

Gene name	Consensus Protein SEQ ID NO:	Representative gene sequences Nucleotide/Protein	Annotation	Sensitivity	Specificity	p-value
		111				
group_55945	22	SEQ ID NO: 67/SEQ ID NO: 112	hypothetical protein	26.37	90.91	0.039
livF_6	23	SEQ ID NO: 68/SEQ ID NO: 113	hypothetical protein	30.77	84.85	0.039
hdhA	24	SEQ ID NO: 69/SEQ ID NO: 114	7-alpha-hydroxysteroid dehydrogenase	46.15	81.82	0.040
ssuA_7	25	SEQ ID NO: 70/SEQ ID NO: 115	Putative aliphatic sulfonates-binding protein	31.87	84.85	0.041
uvrA_2	26	SEQ ID NO: 71/SEQ ID NO: 116	UvrABC system protein A	30.77	87.88	0.049
group_19285	27	SEQ ID NO: 72/SEQ ID NO: 117	hypothetical protein	48.35	83.33	0.051

[0079] Sequences of consensus proteins SEQ ID NO:1 through SEQ ID NO:27 are provided below. In the consensus sequences, X can be any amino acid residue or can be absent.

[0080] **SEQ ID NO:** 1

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXMXXXXXXXXXXX
XXXXXXXXXXXXXXXXXRRRXGXXMLLGTLAAGXXXXRPAXAEPVVEDVRIVDFD
WVDXARQRXVPARLYWPNTSSLRRXVPLVVFSHGLGQSRTGYSYLGRHWSSHGIAS
LHLQHVGSDSTVWTGNPLALLDRIDXAAXEREAIARARDLRFXLDRLLVQDGGXFG
DRIDXRRIVAAGHSYGANTTLIAAGARVIRDGXPLQXRDPRXXAGIVISAPPFYGERD
LRAVLXAVDIPTLHVTATEDXIQLPGRTXSPXXDRLXVYEAIXTPRKALAVFQGGSX
XXXXXXXXXXXXXXHSIFTDXXXXXRPLTGGXNLNPQVKXATAXGXLAFLDLAFR
GDPEPLXXXXXXXXXXXXRAWSSTWXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXKPILAVAPAGFXXXSX
AQPAVAXRXRDRXXXXXXEPLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX

[0081] **SEQ ID NO:** 2

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXMRK
KAXNRXXXXXXXXFTAXXXXXXXILXVAAXXXXGTXXXXLXLGAAEAXRDQXG
NYAQAEWAQDYASPAGMQVQRETXPILSPQTVAATEQMVERYRDIVNRGGWRPVS
GAERLRVGSKGPAVAAVRQRLIVTGDLDPAAGGSGVYDSYVAAGVKRFQARHGLS
QTGAMSMATQQAMNVPADVRLXQLETNVVRLRSYSGDLGRRFVITNIPAALVETVE
NGQVVTLHAAGVGKIDRQSPIMNTKATQINFNPTWTVPASIVKKDLIPKMQKDPNYL
TDNKIRILSGXXEISPRSVNWNSDEGTRYTYRQDSGADFNSMGIVRINIPNPYGVFMH

DTNTKGVFGDDFRFISSGCVRVQNVREYITWLLKDTPGWXREQVEQAIESGKRVDA
NIAQPVPVYWTYITAWATPDGLVQFRDDIYKRDGVNVPSTIGAPTPVASAEPLXPQT
FEPGDXEEXXXX

[0082] **SEQ ID NO:** 3

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
MTEIDTDAFLADLYALREIGRFRTGVHRPTFSAXDMESRRWLMAKLEXCGLEASIDG
IGNVLGRHRGPGPHLLVGSHIETQNEAGWLDGALGVVAGLALARAGLPVDVVAFA
DEEGHFSGGFLGSRSAIGDLTEAEIDAARNRTDGTPLRAALESAGLAGLPRMRLDPA
RYRGFLELHIEQGTQLESAGLHLGVVSGIVAIWQFQIVFDGNQDHAGGTTMAERRDA
GLSAVRLLAAIDREFPKVCGPRSTWTTGRITLDPGGYSIIPGRAEVAFQFRDVSMPVL
ERMEACLEALVRESNRRERCPATLTALSKAIPAPCDPDLMRALSEAAEQVCPGRWQ
VMPSGAGHDAQNIARILPAAMLFVPSIGGISHHWAEDTSDADLAFGVRALGAAAAR
VLAGXXXXXXXXXXX

[0083] **SEQ ID NO:** 4

XXXMRRPDQPALTAAAHGRVSMPFLPTLAVVAILGVASLYCPTSKPGQPXGARVEX
XXXXXXXXXXXXXAGVSAAPGPTDFAPIAAXVAAPGRPPAVIAFAEQYPLDAXVIA
RTGSLPARPAVAARANAHVAAAGRRACPGRRCPETPRSNTDPMAPARGXAADEAE
DALLPSQAMPFAASVVETLVPAARAVGDAANLVRSSARAVQGTVALAVADCLR

[0084] **SEQ ID NO:** 5

MSNSSEPTILAADQDLALRFWGVRGSTPVSGPQYAEFGGSTPCIEVRCGQRMFIVDA
GSGIYNLGQGHRTDLPQEVDLLFSHLHLDHTAGLPFFKPAVLDPDRVINTYCGNLGG
ESAGPTLDRLFAPPLFPVTLDKLCCTFHHHGFEAGQTLAFPDGTRVATILLNHPQGSV
GYRFEHGGRRLCLISDIEHSDPWPDPELAAFVADADLMVYDGMFTDGEYPTCRGWG
HSTWQKGVELARNAGVKALGIIHLHPAHSDTALRDMEADLQAEMPTAFIARERQSLI
VGAPRAVSGRPSGRPAVAREMRRRIKVA

[0085] **SEQ ID NO:** 6

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXMXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXSVXIXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXLXXXXPXXXMRIVXXXXXXAAXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXAAXXXXXAPXXXXXXAAXXVKAGDLXIETP
WLRATPGGAKVAGGYVRIXNTGSXPDRLTGASIPXAGRGEIHSMSMEGGVMKMAP
VXGGLXIKPGETVELKPGGYHLMFEDLTXAPKAGETVSGTLTFERAGTVPVTFTVAP
IGAXAXPXXXXXXXXXXXXXXXXXXXXXXXXXXXGHQHXHHXXXX

[0086] **SEQ ID NO:** 7

MXXXXXXXXXRSAGDAXSASLMDQIAPARTGRTRRRPHPALVVASVALAIGLWFVI
YAVAGRAIWFW

[0087] **SEQ ID NO:** 8

XXXXXXXXXXXXXXXXXXXXXXVDFTLSXSQTHWLTRVRAFIADAILPAAAAVAA
ERAQSRXPSPTMERLKDKARXEGLWNLFLPPSPEHDTDEYRGAGLTNLDYALCAEE
MGRVGIASEVFNCAAPDTGNMEVLHRYGTXAQKDRWLKPLMAGEIRSAFLMTEPE
VASSDATNISTSIRRDGDHYVIDGRKWWSTGVGDPRCRIAILMGKTDPEAPRHRQQS
QILVPMDTPGVRVERLLPVFGYEDAPKGHGEVVLENVRVPAGNLILGEGRGFEIAQG
RLGPGRIHHCMRTIGAAEVALEAMARRLVSRVAFGKRISEQSVWEQRVAEARIDIEM
TRLLCLKAADMMDKVGNKGAKLEIAMIKVAAPRVALKVIDDAIQAHGGAGVSEDF
GLARMYAHIRTLRLADGPDEVHNRSIARLEFGRYTNARXGXXXXXXX

[0088] **SEQ ID NO:** 9

XXXXXXXXXXXXXXXXXMPRXFRXGXXXRAALLXXXXLXXXATPAAAQXXXXXX
XSAASVTLXELSVXGAAAXXXXXXPXXXEXXGSLTVPSVXRQRAALNXTVGSVAF
VDAXXXQDRYANTLRDVLKDVPGVYVQERYGQELRLSVRGSGIARGFHLRGLELL
QDGIPLNLADGSGDFYQVDPLALRSVEVYKGGNALTFGATTLGGAVNXVTPTAYTA
LAPNILRVDGGSFGTIRENFQMSRIXGPLDXLVNGTLTNSDGFRXHEAQRTQNFNANI
GYRIAPGIETRFYLGXYLTDQKLPGTLTLGQSLSTPTLANPTAITGNQSRKVETERIAN
RTSFLLDVGKLDIDTWAIHKSLYHPIFQVIDQDGWTYGISPHWAGTFDIGGFRNDTIL
GLRAFAGQNSALQFVNVRGQRGAQTLNALQSASNXEAYGENRFWFLPDVALMTGA
KAFSSNRTYSDKGGXPXGNPXXRFADVTYEGVNPKIGLLWQPXPDIQVFGDVTRSR
DVPDFSDLVQQNLLSTTFVPLRAQRAWTYEAGARGRIDRLAWDVTLYRSDLRDELI
NFSTNPGXXLNIPAATFNTPRTVHQGVEAAVTLDLARDLXGXGDGLSVTQIWTHND
FRFVGDPVFGNNRIAGIPXDVLRTVMSYRHPSGFHIAPSLDWVPQGAFADHANTLRV
PGYALLGVEAGIDFANGVSLFVDARNLTDARYVSDIAVVANAXXXXXXXAATAGG
PXALAAFYPGSGRSVFXXXXXXXXXXXXXXXXXXGGIRXXXXXXXXXXXXXXXXX
XXXXXXXXASFXXXXXXXXXXXXXXXXXXXXXXXXX

[0089] **SEQ ID NO:** 10

XXXXXXMSETESGPGSVTDADYRNLAETLPQLAWIAEADGTIVWYNQRWYDYTGT
SLDEMRGWGWRTVHHPDHVAAVTERYRAAITLGRSWEDTFPLRGRDGSYRWFLSK
ALPHRDESGRILRWYGTNTDITVRRAVEERLRHSEQRFRALVDASAAVIWNTDAAG
ELMPPQVRWSTYTGQTEEAYQGWGWLDAVHPDDRGHAADAWAACVEARATYEV
EYRLRRHDGAWRAMEVRGVPVLAEDGSLREWVGTCVDVTERKEAEEAVERARQA
AEAANRAKSQFIANMSHELRTPLSAVIGYSEMLGEELEDIGQAALLPDLRKIEAAARH
LLSLINDVLDISKIEAGRMTASAETFTVADLLRDVTDSTGSLVEKKGNRFVLDAGAA
GEAGLGSMHQDQTKIRQCLLNLIGNAAKFTERGTITLTVRRHREAGADWLSFAVAD
TGIGLTEAQIDRLFERFVQADDSTTRQFGGTGLGLAITRAFCRTMGGDIGVASTPGAG
ATFTIRLPATLRPEDAXPPXXXTEAEAHTPVEPHEEHETVLLVDDDPAARELLQRFLE
REGFHVRSANDGRAGLTLARALKPRAILLDVEMPRMDGWAVLHAVRNDPDLAGTP
VIMTSVVAEQGLGQALGATDYFVKPIDWDRLKGLMERYRPAAPNEARVLVVDDDA
DARERLRRSLGREGWTVDEAENGRIALERVSQARPSLILLDLMMPEMDGFGFLRALR
SRPDGDVPVVVLTAKEVTAAEKESLNRQADRVIAKGSMSLAEIGRQLRVLYARSATE
PVPGQLQGLLDRLAEKDAGEPRXXXXXXXXXX

[0090] **SEQ ID NO:** 11

MMMHVKMIAAAAAVVGGLGLAXSAGAAPLAPAGADTITGPAAVSTVAFGCGPGW
APGPYGRCRPIXXXYRRPRFYGPRCFFRPTPWGXXXXXXXXXXXXXXXPRRVCRXX

[0091] **SEQ ID NO:** 12

MXIETVETAGRRRGADLLVEVLRSEGVRYIFGNPGTTELPLIDALTEAPDIAYILALQE
ATAVAMADGYAQGARRPAFLNLHTAGGLGHXMGGLVNSQVSGTPLVVTAGQQDL
RHALTDPLLMGDLVAIADPVMKWAREVTSPDQIPILLRRAFHDXGAAPSGPVFLSLP
MDVMEALTAVPAGETSTIDQRAVAGSLDRLAEKLAAIAPGRLALIAGDEIDASDASA
QMVALADLLAAPVYGSSWPAHIPFPTAHPLWAGNLPTRADAIADILGRYDAVFALG
GKSLITVLYSEVSAVPPGVQVFQLSADVRDLGRTYATXLSTVGDIRASLDALLPLLAP
RLADRADAFAXLRAQAVTARAERRAKLAAAADAAFEDPVIAPLVAAREVARAVGA
ETTIVDEAPATLTHLRTFLDSPSAHQYAAMRGGVLGWGMPAAVGFSLGLDRAPVVC
VVGDGAAMYSPQALWTAAHEKLPVTFVVINNAEYNILKTFMKGQAHYASVRANRF
IAMDLTDPRIDFPALAASMGVPARRVTRAADIAPAIEAGIRSGGANLVEVVVRAT

[0092] **SEQ ID NO:** 13

MLTVTRTPLRISLFGGGTDYPEYFERAPGAVVGAAIDKYIIIAALDLIGCQDYNYRLS
YSRVEHCNNIXEIEHPVVREVLKHFDVNRRLDMSIISDLPAAGSGLGSSSAFTVGFLR

TIYAILNQKPTKIELSKKAIEVEREILRENVGVQDQLHAAFGGINRFDFSGSAIRISPVQ
MSSAAIQQLNASMVLVHTGIARRATTTVAAQIAVTRARAIDKELSELYRLVEECVSL
LEAGTSGWLXQLGEMLSASWRIKRTLSREVSNAVLDDLFEAIIASGAYGAKLCGAG
GGGFFLALIDPDRLPALXERVAPLSVVPIGIDVDGSTLIYRQTRXXXXXX

[0093] **SEQ ID NO:** 14

MSLRALRSSLVASASVLAGTVLPXAVQAQQSVTLGEISVVSTSPVGSGGGNSSXQAQ
IFNGPGPVPPAGSLGPSXXGVRLRGSEQPLYKIPSTVESVTASDITIDRASDNLTTTLAR
RTPGINVSDSQGNNNRVDITYRGFTASPVQGVPQGLAVYQNGVRINEAFGDIVNFDLI
PPQAIQRIDVVTGNPVFGLNALGGAVNIQMKNGFTWQGTEISAWGGSDARTAGYLE
YGKVSGPWSVYFTGDGLNDRGWRYESPSTIGRLYGDIGYRSQDSEFHLIGLAARSFF
GAAAATPVDFTHRDPRAIFTYPQTTTTEVGTLQLTGRVDISPTWDLAGNAYFRRFSQ
TYVDGNDGNFENCSTRSSFRGNLCFEDDGFSPAAGQSQLAFRNQFLILGQXXQNQRI
PFRAGIPYGTLDTTRTEATGFGGSLQAANRDRIFGLSNTFVVGGSIDAANYXFKSSST
LGVINPDLSITTDPSNPFYGNIPGLGTPQLRTAGALGIAPSSVNGSNLYMGLYTLDTLD
VTDRLSLTAGXXXXARLNFARIQSEDLTGFSPDVTGTHYFNKINPVAGLTYRFFDAL
NLYGSYSESNRAPTPLELACANPDRPCLLPNSLVADPPLKQVTGRTYEVGFRGQLPN
TYDGGIITYKIGAFRTDLANDILSLATPGNTARAYFVNVPSTQRQGIEVGGEYXXTAD
YLRVYANYALVDATFQFNGTLSSPNNPLAXXXXXXDDGAIQVRKGNVVPLVPTHQ
XKAGFDYFVTPNWQFGLYLQAFSSSYFRGDESNLNRKLPPYYVLNFQTKYQVTKNL
EVFGLITNLTNNRYXTFGTFAEPGAVAGNLRISDPRTTTLAQPFSVYAGIRYAFGADP
VPMSXPEPIIRKY

[0094] **SEQ ID NO:** 15

XXXXXXXXXXXXMLEIDRLDAWYGPSHVLHGLSLEVRPGEILALVGRNGAGKTTT
MKAVMGLIPKVAGSVRFLGEDLLGRPXHARFPXGLAYVPEDRRIVPGLTVRENLKL
GLLRAXXXXXXXXXXXXXXXXXXAXIKEGPAIAEIAETFPRLAERLDQTAVTMSGGE
QQMLAIARAMIARPKMILLDEPSEGIMPVLVEEMGRLFVAMRXRGVTLLLVEQNVE
WALNLADRAXIIDQGAVVHXSXAAALRADXEIQDRYCAV

[0095] **SEQ ID NO:** 16

MNFVHRYLGESRAVSDADATRLGFVPDTLREPTYFAGTVARHLPFREAISALHHVVV
SDLRFKPKDRTAYFAWLQAHEQELLAEALAEKDSLRAEIEALRAESRDIAARSDAVM
RSFYDARKRYFDYLYRENLDAWIVLDPVITVHPDEIFFEAFSLDESSYGRLSCDHDTF
ARIGDMACGTTNIDYSHALYDEFQKIRSYRDTELAIDPTGFAVQTSGEAAYREDKIDL
PDSWMRGFLQVSSAMAQPAHVVDLHPVDMHAILTRLAARRERHGPRSLRFLLEPGR
PVSVLIEPWNERLTFRRSIYRGGEXAEIRLWGRRRLAILARTLPLARSVRLHLLGTGLP
SFAVVDFGGLRFTLGLSGWTANDWSRAGQFDLLAPRADVDADTAARVFAALRRHH
AADTGQLAAETGLDRSTVEAALGGYVQAGRAMFDLDKRVYRLRELTREPLAPGAL
RFASEQEAKADRFLAAGLVTLGPVEQAGDRRRLSGTVLDDGRSLTPAVELDSDDRM
VGGSCQCGFYTHNRLTRGPCEHMLAVRRLVHAQVEGKPVRWQA

[0096] **SEQ ID NO:** 17

MADPAPITGAPEPVKAAPGAFLTPLFTDRRVAAVLGLGFAQGIPFLLVYATQSAWLV
QAKVPLATIGLMSELTIAYKLKFLWAPFLDRHDAPLIGRWLGRRRGWIVATQILVAL
ALAGVAFGDPAHWLAWTVAFSLALGVAGATQDVVIDGWRITAAPPEQQALMSSWA
EIGFRIGNLAAGAGALYLSDAYGWRVAYLCMAALMAPGTVAALLAPEPPVPETPAT
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGGFVETVWAPIRDLLAXXXRLG
PLALPVLALVAGFRMPGYVSNAMAIPLFKTLGYTNTDIATVTKLFGFWIALGGTFLA
SAIIPRIGMMASLLIGTVTASASHLALAYLAWHGGHGGAAFWTFALTVGIDGFAYAF
ASIVLITYMSRLSATAHAASQYALLTSLCALPGSLLAGFSGFVIEWTGFAWFFVGTSLI
GVPVALLCLLVARRHGPMEPAADAAGDAPNRAT

[0097] **SEQ ID NO:** 18

MSDPSPAVSRAFVLAAGLGKRMRPVTATVPKPLVEVAGKALLDHALDRVAEAGIGT
AIVNVHYLXDLIEGHLARRAEGXAXGPATTVSDERAALLETGGGIRKALPLLGDAPF
VVLNSDSFWLEGPAXNLRRLIDTWDGDRMDGLLLVAPTATSLGYEGAGDFVMDPD
GRLERRGERAVAPFIYAGVAILTPGLFADTPEGAFSLNLLFDRAIAAGRLFGMRLDGQ
WLHVGTPDAIRAAEERVRASARAP

[0098] **SEQ ID NO:** 19

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXMTVRTIIXXXGRDPAAAETSPQIHIXAEEIRRRRREQGLSLEXL
AARSGVSRSMISKIERSEAVPSTVVLSRLAEALGVTFSRLMAPATEREVLLIPASRQPI
LRDEASGYLRRCISPVLPGRGIDWVLNTLPPGASTGEFTAHRRGVSEYIYVLRGRLRA
VIGERAVVMEXGDSLYFEADAGHAFTNVGTEACEYFLVIDPSRVRXXXX

[0099] **SEQ ID NO:** 20

MSDDEKPGWFGRLFGRKXXXXGAPESKSXEPVPAEDXXXXXXXEEXPXXXXXADE
SVTPDPLSPXASESEGQPXFTTAADDVAHVPPPVSEPAPDDPDKNRIPDELEGADLQP
XXXXEPQPPSGXXXXEAPQPDEXXXXXXXXXXESXXEXXAXXXEXAEKRNWWSRL
TGGEXEXXXXXXXPAPAXXXXXXXXXXXXXXXXXXXXXXXAESEPPAEVDIQPV
DSAAALASDAVSGAVEGSEKQGWWSRLTAGMRRTSSALSDRVTGLFTKRKLDATT
LEDLEDALIQADFGVETATRMSEAVGKGRYEKGISPDEVRAILATEIERALEPVALPIE
IDSAKKPYVILTVGVNGAGKTTTIGKLSLKFKAEGRSVMLAAGDTFRAAAIEQLRVW
GDRTGTPVISRAQGSDAAGLAFDAFKEARENGTDVLLIDTAGRLQNKAGLMAELEKI
VRVIRKLDPEAPHATLLVLDATVGQNALSQVELFSQAAPVSGLVMTKLDGTARGGIL
VALATKFGLPVHFIGVGEGVEDLEPFAARDFARAITGLPKEXXX

[0100] **SEQ ID NO:** 21

XXXXXXXMRASLARLMPXXXXXXXXXSTLGRDLPASFVVFLVAMPLCMGIAMAS
GVPAERGLITGIIGGIVVGFLAGSPLQVSGPAAGLAVIVFEFVREHGIDALGPVLVAAG
AIQLLAGALRVGGWFRAISPAVVHGMLAGIGILIVLAQIHVLTDALPKASGLDNLVAI
PAAFFNFVSGPDGNRPGAVIVGLATIVAMIGWEKIRPAKLKLIPGALMGVLAGTLVA
VVGSMDVKRVEVPENIFSAVTMPGXGDWSRLAEPAMILMAVTLAVIASAESLLSAA
AVDRMHDGPRTQYNRELGAQGIGNILCGLAGGLPMTGVIVRSSANVQAGAATRAST
ILHGSWILAFLLVLPMVLKVVPTASLAGILVVTGWRLVSPAHAFHLHERYGLPTAAI
WLATMVMVVATDLLTGVLTGLALSLLQVIPHXYLRGPLKIEGGAAQTVQGGAIQAV
PELRLSGSATFLQLPHLNAALEXRTPEGSPVRLAAQDLRHVDHTCLEMIREWATRRA
KTGSRVEVVGGGASGLHHSXLAMVAHAAXXXPKD

[0101] **SEQ ID NO:** 22

MIVLIAAAIVSGLATATILAPVSALAALIIAPLAASASAILACIFIAWRNTRDDVGPPDL
ETQADAMVAVLCEVAQQGKIVPVAAPVRVRGHRSAXXXX

[0102] **SEQ ID NO** 23

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXMDITGEYRI
XAPRAAVWAALNXXXXXXXXXXXXXXDPEVLAXXXXXXXXXXXXXXXXXXXXRCIP
GCKELTQASPEEMAAKVALKVGPVSATFAGTVXRFEDIRXPEGYTLVGQGNGGMA
GFAKXXXGRAXVSLREEGXETVLTYEAKAEVGGKIASLGGRXXXXXLIQXTXXXSR
KLADQFFGXFAAELGAPXPXAAAXAAXXXXXXXXXXXXXX

[0103] **SEQ ID NO:** 24

XXXXXXXXXXMSLFDLTGKXALITGSSRGIGRAIALRMAEHGARVVISSRKREACEA
VVAEIEAAHGAGRAVAIPASISVKXELEXLVXETESRLGPVDVLVCNAASNPYYGPL
AGISDAQFRKILENNVLSNHWLIQMVAPGMVARRDGAIVIVSSIGALKGSPVIGAYN
VSKAADLQLARNYAVEYGXXXXXPANVRVNCLCPGLIRTDFARALWEDPEMLAAT
TDAAPLRRIGEPDEIAGAAVFLASAAGRFVTGQALVIDGGVTIAXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXX

GKPVTSACTAPAFETHGLAGSISVVDRTYHYVYTDVLPEDCGLAPEKRRTGLFLRTA
QDLSGPKVWSTARKLAGPLPPGTLVRVARAKGMQRWAVSYTCQRPANAPGGPVAD
ICLQYTADLNLDGIGALKLYADPVEAGRSXAYLGLRSGGDGSGRYDRSAHFWMTD
AEGNLDTPAIYPNKAGFLTWLDRLAPTASGRDASSLYGRPVYWATWTVRXIGAQ

Example 4 Soybean Root and Shoot Colonization Efficiency Evaluation

[0107] Six *Methylobacterium* strains were tested for their ability to efficiently colonize soybean roots as follows. Soybean seeds were treated with *Methylobacterium* strains at a target seed titer of 10^5 CFU/seed. Flo-Rite 1706 polymer was used to stick microbe to seed. Seeds were planted in unamended field soil using a randomized complete block design. Plants were grown for 14 days in a greenhouse at 25^0C with regular watering and no fertilizer, after which, roots were harvested. During harvest, the bulk soil (soil that is not physically adhered to the root surface) was shaken off. The root system was then placed in 15 mL of a 0.9% saline solution in a 50 mL conical tube and vortexed for 5 minutes to remove surface-adhered soil. The root system was then removed and placed in a fresh 50 mL tube containing 15 mL of 0.9% saline solution, vortexed for 15 minutes and placed in an ultrasonic bath for 10 minutes. The roots were then removed and the tube contents were poured into the previous tube containing soil. The soil and *Methylobacterium* were centrifuged into a pellet, the supernatant removed, and the soil pellet allowed to dry completely over a period of 96 hours.

[0108] The dry soil was sampled to determine the colonization density of *Methylobacterium* using qPCR primers specifically validated for the target strains using a soil PCR kit. Colonization density was recorded as copy number per gram of tissue.

[0109] The same six *Methylobacterium* strains were tested for their ability to efficiently colonize soybean shoots. Soybean seeds were treated with *Methylobacterium* strains at a target seed titer of 10^5 CFU/seed as described in Example 2 above. Colonization density of *Methylobacterium* on the shoots was determined using qPCR primers specifically validated for the tested strains and recorded as copy number per gram of tissue.

[0110] Results demonstrate that qPCR using specific probes can be used to assess colonization density in a colonization screen and differentiate between efficient and poor colonizers. Results also demonstrate that *Methylobacterium* strains identified as efficient colonizers of soybean shoots also substantially colonize soy root surfaces.

Example 5 Corn Root Colonization Efficiency Screens

Root and Rhizosphere Screen using PCR Quantification

[0111] *Methylobacterium* strains are tested for their ability to efficiently colonize corn roots or associated rhizosphere as follows. Corn seeds are treated with *Methylobacterium* strains at a target seed titer of 10^5 CFU/seed. Flo-Rite 1706 polymer is used to stick microbe to seed. Each experiment includes multiple *Methylobacterium* strains, plus an untreated control treatment (UTC). A strain shown in the past to have limited ability to colonize corn roots may be included as an additional negative control. Corn seeds are planted in unamended field soil, using a randomized complete block design. Plants are grown for 14 days in a greenhouse at 25^0C with regular watering and no fertilizer, after which, roots are harvested. During harvest, the bulk soil (soil that is not physically adhered to the root surface) is shaken off. The root system is then placed in 15 mL of a 0.9% saline solution in a 50 mL conical tube and vortexed for 5 minutes to remove surface-adhered soil. The root system is then removed and placed in a fresh 50 mL tube containing 15 mL of 0.9% saline solution, vortexed for 15 minutes and placed in an ultrasonic bath for 10 minutes. The roots are then removed, and the tube contents poured into the previous tube containing soil. The soil and *Methylobacterium* are centrifuged into a pellet, the supernatant removed, and the soil pellet allowed to dry completely over a period of 96 hours.

[0112] The dry soil is sampled to determine the quantity of *Methylobacterium* using qPCR primers specifically validated for the target strains using a soil PCR kit. Colonization density is recorded as copy number per gram of tissue.

Corn Root Colonization Assay using Non-soil Media

Methylobacterium strains were tested for their ability to efficiently colonize corn roots on corn seedlings grown in non-soil media. The experiment included multiple *Methylobacterium* strains, plus an untreated control treatment (UTC). Corn seeds were planted in calcined clay media, Turface, using a randomized complete block design. The seed was inoculated by micropipette with a 100 microliter suspension of the M-troph isolate in culture media at a concentration of 10^5 CFU/seed. The untreated control received 100 microliters of culture media only.

[0113] Plants were grown for 14 days in a greenhouse at 25^0C with regular watering and no fertilizer, after which, roots were harvested. During harvest, the calcined clay was removed from the roots by shaking. The root system was then placed in 15 mL of phosphate-buffered saline solution and vortexed and sonicated to remove cells. The root wash was then plated on selective media and the number of *Methylobacterium* colonies, as identified by their pink color on the plates, was determined.

[0114] Strains that were identified as the most efficient colonizers or "hits" (as defined by *p*-value in the screen described above) and the poorest colonizers, "non-hits", were analyzed to identify genes associated with enhanced colonization as described above. Gene elements that are positively correlated with *Methylobacterium* colonization of corn roots grown in non-soil media are shown in Table 3 below.

Table 3

Gene name	Consensus Protein SEQ ID NO:	Representative gene sequences Nucleotide/Protein	Annotation	Sensitivity	Specificity	p-value
group_13482	28	SEQ ID NO: 73/SEQ ID NO: 118	hypothetical protein	35.29	95.24	0.029
gpmA_2	29	SEQ ID NO: 74/SEQ ID NO: 119	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase	47.06	90.48	0.038
ecfG_1	30	SEQ ID NO: 75/SEQ ID NO: 120	ECF RNA polymerase sigma factor EcfG	47.06	90.48	0.052
group_47285	31	SEQ ID NO: 76/SEQ ID NO: 121	hypothetical protein	29.41	100.00	0.069
group_59039	32	SEQ ID NO: 77/SEQ ID NO: 122	hypothetical protein	29.41	100.00	0.069
group_35386	33	SEQ ID NO: 78/SEQ ID NO: 123	hypothetical protein	41.18	95.24	0.081
adh	34	SEQ ID NO: 79/SEQ ID NO: 124	putative zinc-binding alcohol dehydrogenase	47.06	90.48	0.082
group_53489	35	SEQ ID NO: 80/SEQ ID NO: 125	hypothetical protein	47.06	90.48	0.082
lgt	36	SEQ ID NO: 81/SEQ ID NO: 126	Prolipoprotein diacylglyceryl transferase	47.06	85.71	0.082
yfiH	37	SEQ ID NO: 82/SEQ ID NO: 127	Laccase domain protein	47.06	85.71	0.082
group_8391	38	SEQ ID NO: 83/SEQ ID NO: 128	hypothetical protein	47.06	85.71	0.082
cyaA__4	39	SEQ ID NO: 84/SEQ ID NO: 129	Adenylate cyclase 1	47.06	85.71	0.082
Vgb_3	40	SEQ ID NO: 85/SEQ ID NO: 130	Virginiamycin B lyase	47.06	85.71	0.082
Group_52947	41	SEQ ID NO: 86/SEQ ID NO: 131	hypothetical protein	47.06	85.71	0.082
pimB_2	42	SEQ ID NO: 87/SEQ ID NO: 132	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase	47.06	95.24	0.084
bmr3_2	43	SEQ ID NO: 88/SEQ ID NO: 133	Multidrug resistance protein 3	29.41	100.00	0.086

Gene name	Consensus Protein SEQ ID NO:	Representative gene sequences Nucleotide/Protein	Annotation	Sensitivity	Specificity	p-value
fabD_1	44	SEQ ID NO: 89/SEQ ID NO: 134	Malonyl CoA-acyl carrier protein transacylase	47.06	85.71	0.093
group_12146	45	SEQ ID NO: 90/SEQ ID NO: 135	hypothetical protein	35.29	100.00	0.097

[0115] Sequences of consensus proteins SEQ ID NO:28 through SEQ ID NO:45 are provided below. In the consensus sequences, X can be any amino acid residue or can be absent.

[0116] SEQ ID NO:28
XXXXXXXXXXXXXXXXLXDXXXGRVYVDCXSCKXXXXXXXXRXGRYXVXSLXERF
GXDISTLXXDLLRXLTASCRYQRXPGAPPARKYEHLCLAAITLPXXXKXXXPVPPGV
PYTIEVWXEXGGXIELHLAXIYPLXMAXAAFEXACXXWPXHEVTLRDRXRIVXKRE
RPXXXATXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXX

[0117] SEQ ID NO:29
XXXXXXXXXXXXXXXXXXVIHDXKXPVTRXLVLVRHGQSVANRSGLFTGLLDSPLT
EQGRXEAVAAGRRLAERSWRFSDAFTSTLTRAVVSGRLILDTLGQPGLIPQRFAALD
ERDYGDLSGLDKTAADARWGAERIETWRRSYAEAPPNGESLRDTVARIVPCYLRSIL
PAVMGGDVLVVAHGNCLRALVMALDDLSPAEVEXLELATGSVRIYEFAADTTIEAR
WIDG

[0118] SEQ ID NO:30
LDEIAPLIEPQIPALRRYAVALLRDREAADDLVQDTLERALSAWSGRRRDGDLRAWL
FTIERNLFLAAVRRRGRRGADLGAEALEQVPDPSADPGAALGARDVLAGLDTLPEEQ
RSVLLLVAVEDLSYAEAAQVLGVPLGTVMSRLSRARTRMRGFVETGRTGLLRRVK

[0119] SEQ ID NO:31
MATSPIWHVVSALAVAGTLAVHGGPAHAAPRQATAGALHFDDGGESYRGYRIEMA
QDVPNAEIGQLRQAAEHQVDIVEATSLDEGTKAFLRRFPVVVXSGSGEXSHYSGGDX
VTIAVEDPKDDRPILLHEYMHVYHFRKMPGGRNNPDILTFYGRAKDGGFYPAGAYL
LKNQGEFFAMTASVYLHGRLAREPFTRDELRQKQPVYYRFLTRLFGPVGA

[0120] SEQ ID NO:32
MTVRNIAACVLLLCTGPALADTLPLRHGAYVSVGTDCKDPPNVALRTYDGGGLGSS
KANDCRSRVLSKQGNVFEIEQDCRQFGGPKVERATERSTIRVDGPDRYTDLTXGXGE
SFRLCPGLKP

[0121] SEQ ID NO:33
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
MALQXXADXXAAXLDAXEQXAAXLLDGXXQGLGALGEAAGQGFAXGVHQGAEX

XGXALEXXXXXLGAXVDQXGERXLAXXERXGQXRXVGLDLXXMGXRPXVEGGX
QXXXALVHEAXEGLVAVVEGXGQGVGALXXQXREXXXAAVDXLGEGIXAGVDXA
GQXLXXXXXXXXXXXXXAALXDXVEVXQXGVEXLXQGXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAX
XXLGXXXQXXAAXLQEPLDXRAXLLEGXXQGXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXREAGVQQVXHXGGRRVEXXAX
XLGXGXDXVGDVXAXXVHEVAQAXLGLLEQGXQALXAXVEGXGELALAGLDQXR
XGGGXLLQXXAQGLXAXVXXGGEAAAAAXDXXVEGXLGLLEQXAQXLGXGXEG
VGEAVAXAVDQXLQXAEXXVEQXAEXXGAXVQXAGGXXERAXQXXDEXXXXAV
QXLXDXAXAXVQGXXQGAGVXVQHDADLVGXVVQQXVXXAHALVEQAADVXX
AARQXVAQAALAXIEHPLGVXDGLLELGEXALAXLVDEXGERGAAXVEQAGELGD
XXADXXEHXXALLVELXRHXXXALVDEAXEQVGAXGEGGXHXLGAGXQXXVGX
XGALVHQXXXXLRAXRQXGLEXAGXLLDGAGEGXDALGHGXVEAAGXXXDXXH
EVAGXXGQLXVXXAGARLDQADRLEGLXAEALVHEAGALVQXXXDXLQAAXDX
XVEXXGAXLDXAGDLAXXAXDVXAEXXGAXXXXXXXXXXXXXXXXXXXXXXX
LXXLHQAAXAXXXDPGHLXXXGXQXLVQGXGAXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXADMX
VDLADQXGHGXGXXGGXLXXXXXXXGXAXLQXAEMGLDQAAGLAGXXRDXLGH
GLAXXGEEXXXLLQXXAXGLAEGGGVAADPVGQDVAXGDGLLQGGEAAAQXLVD
XXDPXVGGLGEXXAXXAEXVGDAXALARHGEHDVAAGAXDXGLGLVXVAGEQL
GXXAAMXAEGXXXLADXXXEVGRDXXGGXXEXAGXXAGXGGEFXVEXXEGXAX
XRDGGXAXXAEXXXXXVGGVAEXGXEDXGXXLDGXXDLXXGGAQXXQXXGEXX
XGIXXXXLGXLAXLXGGXAXXXGXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLXGAAVDLARQXLDX
XXQXXXXXXGXXLEAXAQGREGXLQAXXXXXXPXXDXAAGXLGRRXDXVLQXR
XXDXXGLVEAXXDAVHPLGEXAAAXXXGIXDPLGGXDDVVAHXXGAGXQXXXX
AXXRXXQGXLXXXXAXXDXXXDXXAXXLDLGAQXXAXAXXVLXEGRAXLGEAX
GDLGAXXXEALXXXXXXXXXXXGQXXAPXQEGXXQXRXGALEALXHRXXLXXX
XXADAXXGLXQGXXDXAXXLGQXXGQGXXXXLEXAGXXXGAILEXAXQXXADX
XEAXAHXXAGXRQXVDQXXAAXXHQXXXXVAXLAQXGGDXXAXXLQGAGHAL
AXXXXGRDDPVGXXIQXLXQXXVRAGDGAAHPLGVGDDXLALXXQLVXEGADAD
LVVGIGXLQGGDLAAHQGXQLAXPGEXXLDAVAHRRXLAAXGLRHGQXXVXXEA
LXLGQAHRDLAHXAGDEAHLLGAHXQHXXDXEQXHRXEQRRXXXXXLEXXEAXX
DXXQVAXXLRPXDQXEXXDPXXAGXXGDXVGLXRGAHPQGLLQDADVAPVVVG
HQRAVGRQEAALXAGXVQXXXXQVXGEXRVXXXXXXXXXXXXXDVEAXRXXXXG
GVXXAVEQRIGVXDIXQXXLXXXXXAVXXQGVEXQXLLDXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

[0122] SEQ ID NO:34

MLAMNYRGPYRVRVDRKPVPVCEHVEDAVIRVTRTCICGSDLHLYHGMVPDTRVG
MTFGHEFTGVVEEVGSGVRNLAVGDHVLVPFNIACGRCHFCKQGLFGNCHEANAQ
ATAVGGIYGYSHTAGGYDGGQAEYVRVPYADVGPCKIPXTMDLDDAVLLTDVVPT
GYQAAEMGGIQRGDTVVVFGAGPVGIMAARCAWLFGAGRVIVIDHVEYRLDFARR
YCPAEVYNFRSIGDPVVFLKKTTDSLGADVCIDAVGGDAAGSALHTLMGTKLKLEG

GSAXALHWAINSVKKGGIVSIVGVYGPTGNLIPIGNVLNKGITIRANQASVKRLLPKLI
XHIEAGRLNPKELITHKVPLEEAADAYHLFSAKLDGCIKPVLVPPTARXXXXXX

[0123] SEQ ID NO:35

XXXXMRNRPIDPSTVPGWGVDADPQNDPTYPMRDXARDDXXXXXXXXXXRGMN
WXRPPXXXXQQRARVEVLTSIEYNXRPAVXGTSTPPRGVSGVIRRQAFRYSESQWA
HWLMLMAADRVNVVEGVVDDLXRGRVPNVPAEMGARAELAHNRSGLAXKLALT
GAAIGLGILVSRLARAQRRXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

[0124] SEQ ID NO:36

XXXXXXXXXXXXXXMPPLLALPFPAIDPVAVSIGPIEIKWYALAYIAGLVGGWYYA
RRLVXADSLWGVVKRPTLXDIDDLIVWVALGVVLGGRIGYVLFYNLPMYLXHPXEI
LAIRNGGMSFHGGFLGAILAXLLFARSRGLNGYTMLDIXAVVVPIGLFFGRIANFVNG
ELWGRAAPDFPYAVVFPTGGPVPRHPSQLYEAATEGLLLFIVMALAVRRFGFRKPGL
LGGIFVLGYALARTFCEFFREPDRQLGFLFGXDVXXMGGGVTMGMLLCVPMXLVG
LXYIVLAARGXTRPRKTXXEIXAXEAXQXXXXXX

[0125] SEQ ID NO:37

XXXXXXXXXXXXXXMFIEAPELSSHSXIRHAFFTRQGGVSEGLYASLNGGLGSNDDPE
RVAENRARMCAQLGLPRDXNLVSLYQVHSAEVVTVEAPFPXAERPKADAMVTRVP
GLALGIATADCGPILFADPENRVVGAAHAGWKGALXGVIEATVAAMEALGAERRSI
VAVLGPTIAQASYEVGPDFVERFRAEXPGXERFFXXGXXRPGHAQFDLPGFILXRLA
EAGIGEATALGLCTYADPERFYSFRRTTHRNEPDYGRLISAIALTPXXXX

[0126] SEQ ID NO:38

XXXXATPLXAELARXIRXSGPIGIDRYMALCLGHPVHGYYRTRDPXGXXGDFTTAPE
ISQMFGELVGAWAAXVXXXMGAXXXPDXXXLVELGPGRGTLMADALRALRAAXX
GXXAXXEXHLVETSPVLRRLQAXRLAXAXXXXXXXXXXXXXXXXXPTWHDSVDTL
PDGPLIIIANEFFDXLPVRQFVRTXXGWCERRVGLXXXXXPEGXALAFGLXPEPXPXL
XAXXXXXEAPXGAVLTLPSXGLALMRXLARRLXARGGALLAIDYGHXRPGFGDTX
QAVAGHRFADPLAXPGEADLTXHVDFXALARAAAAEGAAIHGPVTQRDFLLXLGL
AXRAERLKARATPDQAQAIDAAVXRLTDPXPRGMGALFKVLXVSXPALGPLPXLPP
XXXXXXXPXXXXXXXXXXXXXXXXXXX

[0127] SEQ ID NO:39

XXXXXXXXXXXXXXXXXXXXMGPGREAVDAAALLDGFSACLSGLGLPLARATTHA
PTLHPSFRWVMRVWHPGTSSLALRRRHGIEGTPTFHGNTVEHVVETRTPLQCRLDGD
GPLPFPVLGELRNEGLTDYLIAPLXAARGRMGAASWATARPGGFTPIEIETLLALVEP
FSLLFEIKALDDMLGAVLSAYVGRDPARQILAGTVRRGDVRLMRAAMMLTDLRGFG
ELSDRQSPDHVVAALNRMFDAIVPAVEAEGGEVLKYIGDGLLAVFDADRDEAEARR
AALRAAEAALDALATLRDGDRAAFEVGVALHVGEVAYGNIGGGDRVDFTAIGRDL
NVLARVERLCKTYDTPLIATDTFLHGLAHALEPLGIVALRGFAERHALFGCRRTAPV
EAPAVXX

[0128] SEQ ID NO:40

MTHIRTSLRAXLAGAALLLAQGQPGSAAGFDGAIKNNALALNAAGTVAAVSNSEES
AVVVYDVAKGTVLRRLDGFVTPRNIVFAPDGARFYVSDSGTGRITVYETATGKEAG
VLAAGPGAFGTVLSADGGKLYVNNEAASTLTVFDTKTMLADAVIPGFAQPRQGVKL
SPDGKTVFVTNFLGDKITLVDTATNKITGEIAGFDKLRAISITKDGKTLFAANSGRNTI
GVVDVAARKVTSEVEVGKDPYGAALTPDGRFVYSGNLKDNSLSVIDTGTLKVVATV
TGLNEPRQAIAFSADNARAYVLNRDLSVAVVDRAKNAVVSTMKP

[0129] SEQ ID NO:41

MSEPAKPVPPDDPRVRLAEDRTVLAAERTFVAWLRTGLAFLGVGLAAQRFLREVLA
VWPLKVLSLTLIXCALASFAGAAWRDRAIRARLAHAEIPMMPRILTVGIAALLIAISG
LAATALLWA

[0130] SEQ ID NO:42

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXMRVAIV
HYWLIGMRGGEKVVEALCDLYPEADIFTHAYAPQSMSPTIRAHRVRTSFIGRLPFATS
RYKSYLPLMPMALEQLDLRGYDLIISSESGPAKGIIPPSDALHICYCHSPMRYVWNMY
HDYRERTGLLTRLLMPPVAHYVRNWDAVSAGRVHEFIANSDTVARRIETYYRRQAK
VIHPPVDTAAFEXXPDGERGDYHLMVGEMVRYKRPELAIQAFNRLGQPLVVIGGGE
MLRELRSMAGPHIKILGPQPFEVLKHHYARCQALIFPGEEDFGIVPVEAMASGRPVVA
FGKGGVTETVIDGVTGTFFHEQSVDALIDAVQRCRAIGVEPERLVRRAADFGVGRFA
DEISRFVDGVLARERLAAPRPXXXXXXXXXXXXXXXXXXXXXXXXXXPREPSRAY
LVQXXXX

[0131] SEQ ID NO:43

VAXXXXXXXSPXAVTXTQXQAQARPXAPQXRLNPXEIRAIVYGLMTAMLLAALDQ
TIVATAMPTIGLDLGDAANLPWIVTAYLLASTAVTPLYGKLSDIHGRRIMLLIAIATFV
LGSLACALAPTMVTLALARGLQGIGGGGLIALSQTILADIMSPKERARYQVVIAGVFV
AASVAGPLLGGLFAQHLHWSLIFWINLPIGILAFALTNANLKRLPRHERRHRLDWPG
AALMVAGSVTLLLALSWGGVRYPWGSAPVLALLAGAAVLGGGFAARLATAAEPLI
PTEVLKDRVVYSATLAACFAMGTFIGLTIYVPIFLEGVIGLSASESGVALVPLMIGTVT
GATLSGRSMLHFRHYKRVPLAMMXLSLACCATIAWEGRALPFWLMEVLFALLSMGI
GTILPLSTIAIQNAVEPHQLGIATAAMNFFRSLGGALIVAAFGTIVLGGAAGGAGXXX
GGHDVESLIRGADPAQLALTFRXVFLAACLGLLGAFTFLALMEERPLRERXSPRMGG
EPXXGXPX

[0132] SEQ ID NO:44

MTLAILLSGQGGQHPXMFDLTADHPAAQXVFSAARPLLGGTDXXXXPRDLARAXX
XXXXXXXXGGDXLHXNRTGQILCCVAALAAWRALXXXXXGRAIVAGYSIGDLAA
WGVAGRXDXADILALAARRAEAMDAAXXXXXSGEGFGLAGIRGLTLDALDDLAXR
HGCHLAIRNAADSGVVGGRRDALDALCRDAXAAGAQRAVXLXVHTPSHTPLLAAA
SEAFRDALAXVXLRRPPXGAPRLLSGLDGTTVFRPEDGLDKLARQISHTIDWAACLE
ACREYGADRVLELGPGHALATMARXALPXARVHALEEFRSVDGVADWLXRPXXX
XXXX

[0133] SEQ ID NO:45

XXXXMSGPAPDXGRLTVIGVRHHSPACAGLVRRTIAALRPAFVLIEGPVDFNPHLPD
LALGHDLPVAIFSFRADAXGSAASYTPFCAFSPEWQALEAGRAVGAQTLFCDLPAW
DPAFGRRANRYADPHGARAEAAERALAAALGVADQDALWDVLAEAAPEAELPARL
DRYFALLRPPGTDDPAEEARERFMGAYAAHALRAAGDRPVVLVCGGWHADAVRR
HAAQADGTRPEPAPPEPDLRTGSYVVPYAYPRLDRFSGYAAGMPAPGYYERVAEAG
LAPAADWAMTAITAALREAGQVVSTADRIAWRVHAEALARLRAHPAILRADLIDAA
LAALVKDALDRPPAWAAGGAAPGHPALAAMLRALTGRREGRLAPGTRQPPLVADV
AERLRAADLEPGPARRSIDLDWAEPGDRARAHLLHRLXLLGLPGIARXEGPDRAEPG
LPRERFTLVRHPHWLGALIEASLWGGTLEMAAAARITARVEAAPDSLAVLTGALSD
ALFAGLTLEGDLLARLSAGIAAAHDXAALGAAGAGIVRLYRFGDAFXAPARPALAR
LCAALAARALFVVEGIRDPRAGLGAIPLLLACRDLFREVGAEVXGLDELRXPFAAML
GRRLADPETPPALAGAALGFRVACGAAGSDPEAALSRLRRFGLPATLGDFLAGLFAL
AREEIAADATLASVEGLVAAWGDDDFLRALPSLRMAFAWFPPRERERIAVAILRRSG
LGEARAEVEALAWMRQRARPADQAEALAREARVAARLARYGLT

Example 6 Ag Chemistry Tolerance Screen

[0134] *Methylobacterium* strains were screened for tolerance to commonly used agricultural chemicals using a plate assay as follows. Agar plates containing AMS-GluPP media plus one of the below listed chemicals were prepared with concentrations calculated

to approximate the amount that each seed would be exposed to in the field at the middle recommended treatment rate. Concentrations of the chemicals in the plates are provided in Table 4 below.

Table 4

Chemistry	**Concentration**	**Treatment used in Field**	**Chemical Rate in Field (Mid-rate)**
ILeVO (Fluopyram)	3.596 mL/L	Seed Treatment	1.0 fl oz/100 lbs seed
Axyl Shield (Metalaxyl)	0.986 mL/L	Seed Treatment	1.58 fl oz/140,000 seeds
Xtendimax (Dicamba)	160.8 uL/L	Foliar Spray	22 fl oz/Acre*
Headline (Pyraclostrobin)	65.8 uL/L	Foliar Spray	9 fl oz/Acre*

*used 10X rate

[0135] Bacterial strains to be tested were grown for 3-5 days at either 25 or 30^{0}C in 96 well plates in AMS-GluPP media. Using a p200 multichannel pipette set to 175uL, cultures were pipetted up and down approximately 10 times to ensure uniform turbidity throughout. Plates were spotted carefully (to avoid puncturing agar) using a p20 multichannel pipette set to 3.2uL and dispensed until the first stop only to prevent excess spray spots on the plates. Three replicate plates were spotted for each of the strains to be tested. Plates were allowed to fully dry and then inverted and incubated for 5-7 days at room temperature or 30^{0}C. Following incubation, plates were scanned using an Epson scanner. Growth of *Methylobacterium* on plates was visually scored using a rating of 0-3, 0 representing no growth and 3 representing full growth. Control plates were used for comparison to the AgChem plates to ensure accuracy. Scores for each of the three reps were averaged. Strains with a score of greater than or equal to 1.66 were identified as tolerant for a given ag chemical. Strains with a score of $\geq$ 1.66 for each of the ag chemicals tested were considered tolerant to agricultural chemicals.

[0136] Having illustrated and described various embodiments, it should be apparent to persons skilled in the art that the disclosure can be modified in arrangement and detail without departing from such principles.

[0137] Although the materials and methods of this disclosure have been described in terms of various embodiments and illustrative examples, it will be apparent to those of skill in the art that variations can be applied to the materials and methods described herein

without departing from the concept, spirit and scope of the disclosure. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the disclosure as defined by the appended claims or otherwise disclosed herein.

What is claimed is:

1. A method for identifying one or more genetic elements correlated with colonization efficiency of a plant-associated microorganism, wherein said method comprises: (i) screening a population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (ii) identifying a first set of strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment or other strains of said population; (iii) identifying a second set of strains in said population that colonize said plant or plant part at a reduced density as compared to other strains of said population, or at a density that is reduced or not significantly different from that of a non-colonizing control treatment; (iv) comparing the sequences of genetic elements in said first set of strains and said second set of strains; and (v) identifying one or more genetic elements that correlate with colonization efficiency.
2. The method of claim 1, wherein said one or more genetic elements positively correlate with colonization efficiency.
3. The method of claim 1, wherein (i) said one or more genetic elements that are positively correlated with colonization efficiency comprise at least one gene selected from the group consisting of bbsG, hdhA, luxQ, bicA, hddC, hddA, fptA, livF, sutR, cdhR, amaB, ssuA, rbn, ftsY, fecA, gpmA_2, ecfG_1, adh, lgt, yfih, cyaA, vgb_3, pimB_2, bmr3_2, and fabD_1; or (ii) said one or more genetic elements that are positively correlated with colonization efficiency encode at least one protein having an amino acid sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 44, and SEQ ID NO: 45 and/or encode at least one protein having at least 50% sequence identity to a protein selected from the group consisting of SEQ ID NO: 91 to SEQ ID NO:134; and SEQ ID NO: 135.
4. The method of claim 1 wherein said plant-associated microorganisms are bacteria.
5. The method of claim 1 wherein said plant-associated microorganisms are fungi.
6. The method of claim 1 wherein the plant-associated microorganisms in said population are from the same taxonomic genus.
7. The method of claim 1 wherein said plant part is a leaf or stem.
8. The method of claim 1 wherein said plant part is a root or tuber.
9. The method of claim 1 wherein said plant part is a seed.
10. The method of claim 1 wherein said plant is a soybean plant or a corn plant.

11. The method of claim 4 wherein said bacterial are pink-pigmented facultative methylotrophic (PPFM) bacteria.

12. The method of claim 1 wherein said plant-associated microorganism is an endophyte.

13. The method of claim 1 wherein said population comprises 10 or more strains.

14. The method of claim 1 wherein the plant-associated microorganisms in said population are desiccation tolerant.

15. The method of claim 1 wherein the plant-associated microorganisms in said population are tolerant to the presence of one or more agricultural chemicals.

16. The method of claim 15 wherein said one or more agricultural chemicals is selected from fluopyram, metalaxyl, dicamba, and pyraclostrobin.

17. A method of selecting a microbial strain capable of efficiently colonizing a plant, plant cell or plant part, wherein said method comprises detecting in the genome of said microbial strain one or more genetic elements that are positively correlated with colonization efficiency.

18. The method of claim 17, wherein said genetic elements that are positively correlated with colonization efficiency are identified by the method of claim 1.

19. The method of claim 17 wherein (i) said one or more genetic elements that are positively correlated with colonization efficiency comprise at least one gene selected from the group consisting of bbsG, hdhA, luxQ, bicA, hddC, hddA, fptA, livF, sutR, cdhR, amaB, ssuA, rbn, ftsY, fecA, gpmA_2, ecfG_1, adh, lgt, yfih, cyaA, vgb_3, pimB_2, bmr3_2, and fabD_1; or (ii) wherein said one or more genetic elements that are positively correlated with colonization efficiency encode at least one protein having an amino acid sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 44, and 45 or encode at least one protein having at least 50% sequence identity to a protein selected from the group consisting of SEQ ID NO: 91 to SEQ ID NO:134, and SEQ ID NO: 135.

20. The method of claim 17 wherein said microbial strain is a bacterium.

21. The method of claim 17 wherein said microbial strain is a fungus.

22. The method of claim 17 wherein said plant part is a leaf or stem.

23. The method of claim 17 wherein said plant part is a root or tuber.

24. The method of claim 17 wherein said plant part is a seed.

25. The method of claim 17 wherein said plant is a soybean plant or a corn plant.

26. The method of claim 19 wherein said bacteria is s pink-pigmented facultative methylotrophic (PPFM) bacteria.

27. The method of claim 17 wherein said microbial strain is an endophyte.

28. The method of claim 17 further comprising the step of identifying said microbial strain as desiccation tolerant.

29. The method of claim 17 further comprising the step of identifying said microbial strain as tolerant to the presence of one or more agricultural chemicals.

30. The method of claim 29 wherein said one or more agricultural chemicals is selected from fluopyram, metalaxyl, dicamba, and pyraclostrobin.

31. A method of selecting a microbial strain capable of efficiently colonizing a plant, plant cell, or plant part, wherein said method comprises detecting the presence of one or more genetic elements in the genome of said microbial strain, wherein said one or more genetic elements (i) comprise a gene selected from the group consisting of bbsG, hdhA, luxQ, bicA, hddC, hddA, fptA, livF, sutR, cdhR, amaB, ssuA, rbn, ftsY, fecA, gpmA_2, ecfG_1, adh, lgt, yfih, cyaA, vgb_3, pimB_2, bmr3_2, and fabD_1, or (ii) encode at least one protein having an amino acid sequence selected from the group consisting of SEQ ID NO: 1 to SEQ ID NO: 44, and SEQ ID NO: 45; or (iii) encode at least one protein having at least 50% sequence identity to a protein selected from the group consisting of SEQ ID NO: 91 to SEQ ID NO: 134, and SEQ ID NO: 135.

32. The method of claim 31 wherein said one or more genetic elements encode a protein having an amino acid sequence with at least 60%, 70%, 75%, 80%, 85%, 90%, or 95% identity over the entire region of SEQ ID NO: 91 to SEQ ID NO:135.

33. The method of claim 31 wherein said one or more genetic elements encode a protein having an amino acid sequence with at least 80% identity over the entire region of a sequence selected from SEQ ID NO: 91 to SEQ ID NO:135.

34. The method of claim 31 wherein said one or more genetic elements encode a protein having an amino acid sequence with at least 90% identity over the entire region of a sequence selected from SEQ ID NO: 91 to SEQ ID NO:135.

35. The method of claim 31 wherein said method comprises detecting the presence of at least two genetic elements selected from the group consisting of genetic elements that encode a protein having at least 50, 60%, 70%, 75%, 80%, 85%, 90%, or 95% percent identity over the entire region any one of SEQ ID NO: 91 to SEQ ID NO:135.

36. The method of claim 31 wherein said method comprises detecting the presence of at least four genetic elements selected from the group consisting of genetic elements that encode a protein having at least 50, 60%, 70%, 75%, 80%, 85%, 90%, or 95% percent identity over the entire region any one of SEQ ID NO: 91 to SEQ ID NO:135.

37. A method of identifying a yield-enhancing plant-associated microorganism, wherein said method comprises: (i) screening a population of plant-associated microorganisms to identify strains having tolerance to desiccation and tolerance to contact with agricultural chemicals, (ii) screening said population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (iii) identifying strains having tolerance to desiccation and contact with agricultural chemicals and strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment; (iii) screening said population of plant-associated microorganisms or said strains identified as having tolerance to desiccation and agricultural chemicals and as having the ability to colonize a plant or plant part to identify strains that efficiently colonize a target plant host; (iv) contacting strains identified as having tolerance to desiccation and agricultural chemicals and that efficiently colonize a plant host with said plant host or a part of said plant host; (v) growing said plant host under field conditions; (vi) determining harvested yield of said plant host; and (vii) identifying strains that improve yield of said plant host.

38. The method of claim 37 wherein said yield-enhancing plant-associated microorganism is a bacterium.

39. The method of claim 37 wherein said yield-enhancing plant-associated microorganism is a fungus.

40. The method of claim 38 wherein said bacterium is a *Methylobacterium*.

41. The method of claim 37 wherein said plant host is a soybean plant.

42. The method of claim 37 wherein said one or more agricultural chemicals is selected from fluopyram, metalaxyl, dicamba, and pyraclostrobin.

43. The method of claim 37 wherein said screening in (iii) comprises comparing strains in a colonization screen to identify strains that colonize said plant or plant part at an enhanced density.

44. The method of claim 37 wherein said screening in (iii) comprises detecting in the genome of said microbial strain one or more genetic elements that are positively correlated with colonization efficiency.

45. The method of claim 31 wherein said one or more genetic elements is detected by a nucleic acid detection technique or a sequence comparison, wherein the genetic element has at least at least 50, 60%, 70%, 75%, 80%, 85%, 90%, or 95% percent identity to a sequence selected from the group consisting of SEQ ID NO: 46 to SEQ ID NO: 89, and SEQ ID NO: 90.

46. The method of claim 31 wherein said one or more genetic elements is detected by direct sequencing of the genome of a target microorganism.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/012041

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search 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 No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet(s).

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 additional fees, this Authority did not invite payment of additional fees.

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.:

1-4, 6-12, 14-20, 22-38, 40-44, and 46 to the extent that they read on SEQ ID NO:91.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/012041

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A01H 5/00; A01N 63/00; A01P 21/00; C12N 1/00; C12N 1/20; C12N 5/04 (2020.01)
CPC - A01N 25/00; C12N 1/00; C12N 1/20; C12N 5/04; C12Q 1/68 (2020.02)

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)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 424/93.4; 435/252.1; 504/117; 504/116.1; 800/298 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EGAMBERDIEVA et al. "Salt tolerant Methylobacterium mesophilicum showed viable colonization abilities in the plant rhizosphere," Saudi J Biol Sci, 04 July 2015 (04.07.2015), Vol. 22, Pgs. 585-590. entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	US 2013/0031673 A1 (GRANDLIC et al) 31 January 2013 (31.01.2013) entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	MADHAIYAN et al. "Leaf-residing Methylobacterium species fix nitrogen and promote biomass and seed production in Jatropha curcas," Biotechnol Biofuels, 21 December 2015 (21.12.2015), Vol. 8, Pgs. 1-14. entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	US 2016/0302423 A1 (NEWLEAF SYMBIOTICS, INC.) 20 October 2016 (20.10.2016) entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	WO 2018/106899 A1 (NEWLEAF SYMBIOTICS, INC.) 14 June 2018 (14.06.2018) entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	US 2016/0046925 A1 (NEWLEAF SYMBIOTICS, INC.) 18 February 2016 (18.02.2016) entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46
A	WO 2015/114552 A1 (UNIVERSITY OF PRETORIA) 06 August 2015 (06.08.2015) entire document	1-4, 6-12, 14-20, 22-38, 40-44, 46

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 April 2020	07 MAY 2020
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Continued from Box No. III Observations where unity of invention is lacking

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

Group I+: claims 1-46 are drawn to one or more genetic elements, and methods comprising the same.

The first invention of Group I+ is restricted to a genetic element, and methods comprising the same, wherein the genetic element is selected to be a cdhR gene of SEQ ID NO:91. It is believed that claims 1-4, 6-12, 14-20, 22-38, 40-44, and 46 read on this first named invention and thus these claims will be searched without fee to the extent that they read on SEQ ID NO:91.

Applicant is invited to elect additional genetic elements, each with specified SEQ ID NO to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a genetic element, and methods comprising the same, wherein the genetic element is selected to be an amaB gene of SEQ ID NO:93. Additional genetic elements will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general 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:

The Groups I+ formulas do not share a significant structural element for enhancing colonization efficiency of a plant-associated microorganism, requiring the selection of alternatives for the cell genetic element, where "said one or more genetic elements that are positively correlated with colonization efficiency encode at least one protein having an amino acid sequence selected from the group consisting of SEQ ID NO:1 to SEQ ID NO:44, and SEQ ID NO:45 and/or encode at least one protein having at least 50% sequence identity to a protein selected from the group consisting of SEQ ID NO:91 to SEQ ID NO:134; and SEQ ID NO:135".

Additionally, even if Groups I+ were considered to share the technical features of a method for identifying one or more genetic elements correlated with colonization efficiency of a plant-associated microorganism, wherein said method comprises: (i) screening a population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (ii) identifying a first set of strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment or other strains of said population; (iii) identifying a second set of strains in said population that colonize said plant or plant part at a reduced density as compared to other strains of said population, or at a density that is reduced or not significantly different from that of a non-colonizing control treatment; (iv) comparing the sequences of genetic elements in said first set of strains and said second set of strains; and (v) identifying one or more genetic elements that correlate with colonization efficiency; a method of selecting a microbial strain capable of efficiently colonizing a plant, plant cell or plant part, wherein said method comprises detecting in the genome of said microbial strain one or more genetic elements that are positively correlated with colonization efficiency; a method of identifying a yield-enhancing plant-associated microorganism, wherein said method comprises: (i) screening a population of plant-associated microorganisms to identify strains having tolerance to desiccation and tolerance to contact with agricultural chemicals, (ii) screening said population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part; (iii) identifying strains having tolerance to desiccation and contact with agricultural chemicals and strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment; (iii) screening said population of plant-associated microorganisms or said strains identified as having tolerance to desiccation and agricultural chemicals and as having the ability to colonize a plant or plant part to identify strains that efficiently colonize a target plant host; (iv) contacting strains identified as having tolerance to desiccation and agricultural chemicals and that efficiently colonize a plant host with said plant host or a part of said plant host; (v) growing said plant host under field conditions; (vi) determining harvested yield of said plant host; and (vii) identifying strains that improve yield of said plant host; these shared technical features do not represent a contribution over the prior art.

Specifically, WO 2015/114552 A1 to University of Pretoria discloses a method for identifying one or more genetic elements correlated with colonization efficiency of a plant-associated microorganism ([t]he invention further relates to procedures for identifying effective plant growth promoting Rhizobacterial strains, Pg. 1, second paragraph), wherein said method comprises: screening a population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or plant part (screening the isolates for efficacy as plant growth enhancers and biocontrol agents, Pg. 1, second paragraph; [t]he purified isolates were subsequently screened for ....root colonization of tomato and wheat seedlings, Pg. 10, last full paragraph); identifying a first set of strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment or other strains of said population ([t]he most effectual isolates were identified based on partial 16S rRNA gene sequences and assessed for MOA including ...root colonization of tomato and wheat seedlings, Pg. 10, last full paragraph; negative control - plants were untreated and uninfected, Pg. 13, second paragraph); identifying one or more genetic elements that correlate with colonization efficiency ([t]he most effectual isolates were identified based on partial 16S rRNA gene sequences and assessed for MOA including ...root colonization of tomato and wheat seedlings, Pg. 10, last full paragraph); a method of selecting a microbial strain capable of efficiently colonizing a plant, plant cell or plant part ([t]he invention further relates to procedures for identifying effective plant growth promoting Rhizobacterial strains, Pg. 1, second paragraph), wherein said method comprises detecting in the genome of said microbial strain one or more genetic elements that are positively correlated with colonization efficiency ([t]he most effectual isolates were identified based on partial 16S rRNA gene sequences and assessed for MOA including ...root colonization of tomato and wheat seedlings, Pg. 10, last full paragraph); a method of identifying a yield-enhancing plant-associated microorganism ([t]he invention further relates to procedures for identifying effective plant growth promoting Rhizobacterial strains, Pg. 1, second paragraph), wherein said method comprises: screening a population of plant-associated microorganisms to identify strains having tolerance to desiccation and tolerance to contact with agricultural chemicals (screening the isolates for efficacy as plant growth enhancers and biocontrol agents, Pg. 1, second paragraph), screening said population of plant-associated microorganisms to determine the ability of strains in said population to colonize a plant or

plant part (screening the isolates for efficacy as plant growth enhancers and biocontrol agents, Pg. 1, second paragraph; [t]he purified isolates were subsequently screened for ....root colonization of tomato and wheat seedlings, Pg. 10, last full paragraph); identifying strains having tolerance to desiccation and contact with agricultural chemicals and strains in said population that colonize said plant or plant part at an enhanced density as compared to a non-colonizing control treatment ([t]he most effectual isolates were identified based on partial 16S rRNA gene sequences, Pg. 10, last full paragraph); contacting strains identified as having tolerance to desiccation and agricultural chemicals and that efficiently colonize a plant host with said plant host or a part of said plant host (applying to plants, plant tubers, plant seeds, or soil surrounding plants or plant tubers or seeds, the biocontrol and biofertilizer composition as described, under conditions that are effective to enhance growth of the plants, Pg. 5, fourth paragraph); growing said plant host under field conditions (effect of rhizobacterial seed treatments on yield of maize in three different soil types under field conditions, Pg. 7, first paragraph); determining harvested yield of said plant host ([e]ffect of rhizobacterial seed treatments on yield of maize in three types of soil under field conditions, expressed as the difference in yield (tons/hectare) compared to the untreated control, Pg. 7, second paragraph); and (vii) identifying strains that improve yield of said plant host (Pseudomonas fluorescens strain N-04 produced significantly higher biomass yields based on total wet mass, wet shoot mass and wet root mass as compared to the positive control, Pg. 14, first paragraph).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.