INTRODUCTION

Increased public awareness about the health benefits of blueberries has elevated its worldwide demand in recent years; thus, researchers and breeders seek ways to make blueberry cultivation more efficient and fruitful. An increased understanding of the blueberry genome and pathways, for example, can facilitate selection for climatic adaptation, enabling blueberry growth in new regions as well as greater cultivation in existing growth regions. Through pathway elucidation, blueberry fruit quality can also be improved, providing consumers and breeders with the traits they desire. For consumers, these characteristics include increased antioxidants and dietary fiber, improved taste, cheaper costs, and added vitamin content. For breeders, disease resistance, pest resistance, increased yields, and increased zone hardiness are qualities of interest. For a beginner in this discipline, a clear understanding of mapping, sequencing, genome assembly, and proteomics will provide the necessary framework to explore and contribute intellectually to the growing field of plant genomics. This paper presents an overview of these topics with an emphasis on how they pertain to the study of blueberries.

OVERVIEW OF BLUEBERRIES

Blueberries provide a perfect case study to explore current research in the fields of genomics, bioinformatics, and proteomics, as it is not one of the highly characterized plant species such as rice (Oryza sativa) or Arabidopsis thaliana. Humans have gathered and consumed blueberries for thousands of years. Blueberries are native to North America and were a staple in the diet of Native Americans and early settlers. Additionally, blueberries were used for medicinal purposes, relieving fevers, headaches, and persistent coughs (Tichace, 2004). Currently, North America is the largest commercial blueberry producer in the world, producing 600 million pounds of blueberries in 2012. Production in North America has been on the rise, with the total supply increasing by an average of 20% every two years since 2008 (Brazelton, 2013).

HEALTH BENEFITS

Consumption of blueberries has increased dramatically in the past 15 years due to consumers’ growing knowledge of its health benefits. Blueberries, along with several other berries, contain various types of anthocyanins, which are compounds with anticarcinogenic and therapeutic properties. These compounds have the ability to negate the effects of free radicals in the body, thereby protecting DNA integrity, improving brain function, preventing cancer cell formation, and reducing cardiovascular disease (Zafra-Stone et al., 2007). In fact, a study conducted in rats correlated the consumption of blueberries with protection against neurodegeneration and cognitive impairment. During an 8-week period, rats were injected with kainic acid (KA) and then fed either a diet containing 2% blueberry extract or a control diet. After analyzing a variety of performance variables, researchers concluded that KA-treated rats showed clear signs of impaired learning performance, but the blueberry diet reduced the impairment (Duffy et al., 2008).

GENETICS AND GROWTH

Blueberries have a varied genetic background. The basic chromosome number (x) is 12 (Rowland and Levi, 1994), and seven different ploidy levels have been found in the wild: 2x, 3x, 4x,
The blueberry genome is estimated to be 508–608 Mb/1 C DNA, which is four to five times larger than the 125-Mb Arabidopsis thaliana genome and several times larger than other fruiting plant genomes such as the strawberry (240 Mb; Shulaev et al., 2011), grape, and raspberry genomes (Die and Rowland, 2013).

In terms of growth, blueberries require significant organic matter and thrive in acidic, sandy soil, which was originally considered marginal for crop production (Trehane, 2004; Rowland et al., 2012a). There are numerous varieties of blueberry in existence around the world, but four of the major varieties are Vaccinium corymbosum (Northern highbush), V. darrowii (Southern highbush), V. angustifolium (lowbush), and V. ashei (rabbit-eye; Trehane, 2004). Different varieties, however, have particular requirements, such as chilling hours. The number of chilling hours is a measure of accumulated hours of temperatures below 7°C in the dormant season (Costa and Quatrano, 2004).

Genetic studies on blueberries have the potential to significantly improve fruit quality and the breeding process. Blueberry breeding will be simplified, for instance, through the use of genetic markers to identify desired seedlings. Rather than waiting for seedlings to mature and then examining their phenotypes, breeders can proactively select and cultivate individuals with desired genotypes, speeding up the screening process. Utilization of genetic markers will make traditional breeding methods more efficient and less random (Hancock et al., 2008).

SEQUENCING AND ASSEMBLY

Sequencing genomes has the potential to solve large-scale and small-scale biological problems by characterizing key genes involved in various biological processes. These include genes whose mutations lead to disease susceptibility (NOD2 and Crohn's disease in humans; Ogura et al., 2001), genes involved in flowering and fruiting (FT and CO in Arabidopsis thaliana; Tureck et al., 2008), and genes controlling anti-inflammatory responses (IL6 in humans; Xing et al., 1998). The recent development of easier and cheaper sequencing technologies has led to a rapid increase in the number of sequenced genomes. With regard to plants, Arabidopsis thaliana was the first sequenced plant genome in 2000 and has become the model organism for plant genomics due to its small genome size, short life cycle, and low chromosome number. These features are extremely desirable in the sequencing process. When sequencing a new plant species, various elements must be considered including the overall genome size, short life cycle, and low chromosome number. These features are extremely desirable in the sequencing process. When sequencing a new plant species, various elements must be considered including the overall genome size, short life cycle, and low chromosome number. These features are extremely desirable in the sequencing process.

ASSEMBLY PROCESS

Using millions of sequenced short reads, genome assembly combines the reads into contiguous segments (contigs), which are ideologically reconstructed into whole chromosomes (Boetzer et al., 2011). Initially, this process compares shotgun sequence reads and overlaps the reads using an indexing technique to identify the sequences that commonly overlap. After indexing, another algorithm aligns similar reads, laying out all of the alignments. To get a final DNA sequence, the layout is condensed down to a consensus (Pop et al., 2004b). Additional algorithms calculate the distance between mate pairs and construct supercontigs, also known as scaffolds, which are multiple contigs joined together (Boetzer et al., 2011). For example, if one end of a mate pair is located in one contig and the opposite end is located in another contig, the distance between the mate pairs ascertainment the distance between these two contigs in the genome. Multiple mate pair links are desired in order to corroborate the correlation between two
contigs. Since the gap between these contigs lacks sequenced data, “N”s fill the empty space (Pop, 2009).

**DE NOVO AND COMPARATIVE ASSEMBLY**

Assembly of contigs and scaffolds utilizes two approaches: de novo assembly and comparative assembly. In de novo assembly, all reads are assembled based on algorithms, and no outside template is used. In comparative assembly, the sequence assembly can be aligned to a reference genome, a BAC (Bacterial Artificial Chromosome) library, or linkage groups. A reference genome utilizes a well-annotated genome that is similar to the species of interest. *Arabidopsis thaliana*, for example, is the primary reference genome for most plant studies. BAC libraries can also be used in comparative assembly. In this, a query genome sequence of several thousand to over one hundred thousand bases is ligated onto bacterial vector DNA. Bacteria take up the DNA and then replicate on an agar plate. Following this replication, the bacterial DNA with the original query sequence is cut with restriction enzymes, and overlapping of different BACs can determine the restriction enzyme locations for a particular sequence (Shizuya et al., 1992; Shizuya and Kouros-Mehr, 2001). BAC libraries are useful because they only include the actual genome without additional cloning artifacts or alterations from the cloning process (Choegogawa et al., 2001). Finally, mapping of linkage groups can assist in assembly. For instance, mapping of restriction enzyme sites can result in an ordered directory of all restriction fragment length polymorphisms for the genome. The location of these restriction enzyme sites can be matched to the assembled contigs, helping to merge the contigs into scaffolds and eventually chromosomes (Pop, 2009).

**ASSEMBLY PROGRAMS**

A variety of programs exist for genome assembly including Arachne (Batzoglou et al., 2002), Bambus (Pop et al., 2004a), Euler (Pevzner et al., 2001), MIRA (Chevreux et al., 1999), Newbler (Margulies et al., 2005), and Velvet (Zhao and Birney, 2008). These programs can be specialized for contig construction, scaffold building, or both, though most use a greedy approach, which starts with the reliable baseline data and then slowly merges additional data as long as this supplemental information agrees with the current sequence construction. Each assembler has unique strengths and weaknesses, as seen by comparing the MIRA and Newbler assemblers. The Newbler assembler, which is distributed by 454 Life Sciences, fail to account for repeats in an organism’s genome, therefore misassembling the genome but creating fewer scaffolds than the MIRA assembler (Lai et al., 2012). The MIRA assembler flags these repeats but results in a larger number of extremely small scaffolds. In addition, different assemblers work with different read types. For example, the Newbler assembler works best with 454 sequences. When deciding on assembling software, scientists must balance the strengths and limitations of each genome assembly program.

**GENOME MAPPING**

Genome mapping is an essential step for assembling highly repetitive genomes. There are two types of mapping: genetic mapping and physical mapping. Genetic mapping approximates the distance between genetic markers by comparing recombination frequencies. Genetic mapping requires cultivation of populations and acquisition of marker data that have been specifically designed for mapping. Physical mapping, on the other hand, anchors a linkage map to physical locations using sequencing, BAC libraries, or restriction enzyme sites (Collard et al., 2003). The end goal of genetic mapping is to create large linkage groups that sufficiently cover the organism’s chromosomes and can be used to assemble the sequence into chromosomes.

**MAPPING POPULATION**

To make a genetic map, the distance between loci must be calculated by creating and genotyping the recombination frequencies between loci within a particular population. Accurately estimating recombination frequencies in a population requires a thorough understanding of the species, particularly the reproductive methods. For example, some plants can reproduce by self-pollination, while others require cross-pollination due to self-incompatibility. Using this knowledge, mapping populations create a population of individuals with stable segregation ratios. The genetic markers are then identified on each individual, and the recombination ratios for each marker are plotted in a matrix. The ratios of recombination determine the centiMorgan linkage distances between markers, such that 1 cm equals 1% recombination (Semagn et al., 2006).

**GENETIC MARKERS**

Genetic markers represent molecular differences between or within a species and are used to tag particular sequences due to their proximity to genes of interest. Genetic markers distinguish polymorphisms between an individual offspring, assess genetic relationships, and assist in linkage map construction (Collard et al., 2003). As for marker types, RFLPs (restriction fragment length polymorphisms) are differences between homologous DNA sequences that are digested by restriction enzymes. Restriction enzymes cut specific locations on DNA, resulting in different length fragments. RAPDs (random amplification of polymorphic DNA) are DNA segments that are randomly amplified by PCR and total 8–12 nucleotides (Levi and Rowland, 1997). Another type of marker, EST-PCR (expressed sequence tag-polymorphic chain reaction), is more commonly used today. ESTs are short DNA fragments of 200–500 nucleotides and are generated from sequencing one or both ends of an expressed gene, called complementary DNA (cDNA). Since EST markers are transcribed from mRNA, they only contain expressed genes and do not include introns (Dhanaraj et al., 2004). In addition to ESTs, SSRs (simple sequence repeats) have regularly been used in plant genomics. SSRs, also called microsatellites, are short repeated DNA sequences of 2–6 base pairs (Boches et al., 2006).

**BLUEBERRY MAPPING**

The first blueberry genetic markers published were RFLPs (Haghighi and Hancock, 1992). This study utilized RFLP segregation in blueberry mitochondrial DNA to distinguish diverse highbush cultivars. Since the 1990s, other genetic markers have been analyzed in blueberry including RAPD (Rowland and Levy, 1994), SSR (Levy and Rowland, 1997), and EST-PCR markers (Dhanaraj et al., 2004). Focusing on EST-PCR...
After a sequence has been declared ready for annotation, repeats are costly and time intensive (Yandell and Ence, 2012). For a less directive annotation, results in a more accurate approach to gene prediction, which can include gene expression using the transcriptome or ESTs, protein-protein interactions, and metabolomics of ESTs and RNA-seq data, enable deeper insight into an organism’s biomolecular pathways (Hiirai and Saito, 2004).

BLUEBERRY PATHWAYS

For blueberries, the anthocyanin and flavonoid biosynthesis pathways are a primary focus of research due to their health and anticarcinogenic properties. A recent study produced 1,000 transcripts and 800 transcription factors relating to antioxidant pathways and discovering the identities of biochemical compounds, can provide insights into the biosynthesis of under-investigated natural products. Specifically, many studies analyze comparative transcriptomics and gene expression to characterize biosynthesis pathways. Following the introduction of stress to an organism, random mutations arise, potentially altering the natural processes and genetic functions. These conditional stress factors often test a plant’s ability to tolerate adverse conditions. By examining the alterations in the organisms, scientists can take a “top-down” or “bottom-up” genetics approach by studying phenotype to DNA or DNA to phenotype respectively (Fahn, 2001). This strategy and others, such as target analysis, profiling, fingerprinting, and metabolomics of ESTs and RNA-seq data, enable fuller elucidation of the metabolic pathways. Future investigations should place focus on a closer relative to the blueberry, such as the grape (Vitis vinifera), and strive for full elucidation of the metabolic network interactions.

BLUEBERRY GENOME

Dr. Allan Brown of North Carolina State University is currently working on a draft genome of diploid V. corymbosum. Dr. Brown’s de novo approach utilizes both Roche 454 and Illumina GAIIx libraries. The working assembly consists of approximately 500 million base pairs, which encode 25,000 genes. Though the genome is not yet published, these recent developments will be valuable for understanding blueberry cultivation, pathways, and nutritional value (Die and Rowland, 2012).

CONCLUSION

As the disciplines of genomics, bioinformatics, and proteomics evolve, clear introductions can quickly become convoluted. For novices interested in these fields, a baseline understanding of key markers, the first blueberry EST-PCR study examined cold acclimation genes using fruit, flower bud, leaf, and stem tissue (Dhawan et al., 2004). More recently, 110 EST-PCR markers have been mapped into 16 genomic linkage groups (Rowland et al., 2012a). EST markers have also been used in evaluating blueberry species including lowbush and rabbiteye (Rowland et al., 2010). Furthermore, EST library development has assisted with finding blueberry SSR markers (Bocher et al., 2005; Bassil, 2012). Continued marker studies will help with identification of blueberry cultivars and management of germplasm in gene banks.

ANNOTATION

ANNOTATION PREPARATION

Following assembly of contigs and scaffolds, various test statistics are used to determine if an assembly is ready for annotation. The most widely used statistic is N50, a scoring metric that describes the length of assembled scaffolds (International Human Genome Sequencing Consortium, 2001). The sequences are sorted according to size and then summed from the largest sequences in decreasing order until half of the total size of the sequences has been tallied. The N50 statistic is the size of the smallest contig or scaffold within this set of the largest sequences. Other statistics, such as percent gaps and percent coverage, ensure that there is not significant missing data (Yandell and Ence, 2012). Though N50 is frequently used, it only describes part of the assembly and has been disputed as an ideal metric for describing assemblies (Baker, 2012). Other metrics based on the N50 have also been proposed (Earl et al., 2011; Mäkinen et al., 2012).

REPEAT MASKING

After a sequence has been declared ready for annotation, repeats must be identified and masked. Repeats are short sequences that occur multiple times throughout a genome. Plant genomes contain a high percentage of repetitive DNA. This repetitive DNA can prevent the computer from gathering evidence and correctly assigning sequence locations, which is especially problematic for sections of coding DNA. Many computer programs can identify and mask repeats, such as RepeatMasker (Tempel, 2012) and RepeatScout (Price et al., 2005), allowing the program to ignore the repeat (Bao and Eddy, 2002). Once the repeats are masked, the process of gene annotation commences.

ANNOTATION PROCESS

There are two types of gene annotation: ab initio and evidence-driven. Ab initio gene prediction uses computer-driven mathematical models to identify putative genes and determine their intron-exon structures. These predictions can be advantageous, as they do not require external evidence, saving time and money. Unfortunately, this approach caps the prediction’s accuracy at 70%. Evidence-driven gene annotations, on the other hand, uses data obtained from further analysis. This additional data, which can include gene expression using the transcriptome or ESTs, protein isolation, or experimental evidence based on cloning and characterization, results in a more accurate approach to gene prediction. This accuracy, however, comes at a price, being more costly and time intensive (Yandell and Ence, 2012). For a less expensive and time-consuming method, automated tools such as BLAST (Basic Local Alignment Search Tool) can query an input of known genes against the unknown genome and locate similarities between the two samples (Altschul et al., 1990). These results must be individually examined and interpreted in light of additional ab initio and evidence-based predictions to obtain the final genome annotation.

PATHWAY ELUCIDATION

Pathway elucidation, which involves building gene expression pathways and discovering the identities of biochemical compounds, can provide insights into the biosynthesis of under-investigated natural products. Specifically, many studies analyze comparative transcriptomics and gene expression to characterize biosynthesis pathways. Following the introduction of stress to an organism, random mutations arise, potentially altering the natural processes and genetic functions. These conditional stress factors often test a plant’s ability to tolerate adverse conditions. By examining the alterations in the organisms, scientists can take a “top-down” or “bottom-up” genetics approach by studying phenotype to DNA or DNA to phenotype respectively (Fahn, 2001). This strategy and others, such as target analysis, profiling, fingerprinting, and metabolomics of ESTs and RNA-seq data, enable fuller elucidation of the metabolic pathways.
concepts is essential. This background will enable these individuals to explore the currently relevant subjects, such as the sequencing of organisms or the elucidation of pathways. In fact, much is still unknown about many currently published genomes. Even the most complete and understood genome, the human genome, is only 10% characterized (Maher, 2012). Additionally, there are many organisms that have not yet been investigated. Examination of new organisms will yield breakthroughs and game-changing discoveries. Of the many organisms with a paucity of research, the blueberry is especially interesting, as scientists and consumers are conscious of its health benefits. Further exploration into the blueberry genome will allow researchers to relate genotype to phenotype, thus providing scientists and farmers with the necessary knowledge to produce blueberries that are more nutritious and desirable.

ACKNOWLEDGMENTS

This work was carried out as part of the Plant Pathways Eucation Project using resources from a collaboration of the Cabarrus Economic Development Corporation, Catawba College, David H. Murdock Research Institute, Dole Nutrition Research Cabarrus Economic Development Corporation, Catawba College, Students’ perspective on genomics 2013| Volume 4| Article 245 |

REFERENCES

Boeck H., and Hancock F. J. E. (1992). DNA transcription fragment length variability in the genome of highbush blueberry. HortScience 27, 46–47.

Hamilton, J. P. and Buell, C. R. (2012). Advances in plant genome sequencing. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica 142, 169–199. doi: 10.1007/s10681-005-1681-5

Bhunjan, A. L., Sovri, J. P. and Rowland, L. J. (2004). Analysis of gene expression associated with cold acclimation in blueberry (Vaccinium uliginosum) using expressed sequence tags. Plant Sci. 166, 863–872. doi: 10.1016/j.plantsci.2003.11.013

Doe J. Y. and Rowland, L. J. (2011). Advent of genomics in blueberry. M. Breeding 12, 493–504. doi: 10.1093/molbev/msr121.5-169-5

Duffy, K. B. Spangler, E. L. Devon, B. D., Geo, Z., Broker, J. L., Jones, A. M., et al. (2008). A blueberry-enriched diet provides cellular protection against oxidative stress and reduces a kainate-induced learning impairment in rats. Neurochem. Aging 29, 1680–1689. doi: 10.1016/j.neurobiolaging.2007.04.002

Earl, D., Abraham, K. M., John, F., Darling, A., Lin, D., Fau, J., et al. (2011). Assemblyathon 1: a competitive assessment of de novo short read assembly methods. Genome Res. 21, 2224–2241. doi: 10.1101/gr.120599.111

Egan, A. N., Schlueter, J. and Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. Annu. Rev. Bot. 63, 175–198. doi: 10.1146/annurev-arbo-032510-102609

Fank, O. (2011). Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp. Funct. Genomics 2, 135–168. doi: 10.1002/qagb

Highkin, B. E. Schlueter, J. and Spooner, D. M. (2012). Applications of next-generation sequencing in plant biology. Annu. Rev. Bot. 63, 175–198. doi: 10.1146/annurev-arbo-032510-102609

Hiram, M. Y. and Saito, K. (2004). Post-genomics approaches for the elucidation of plant adaptive mechanisms to sulphur deficiency. J. Exp. Bot. 55, 1871–1879. doi: 10.1093/jxb/erh184

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860–921. doi: 10.1038/3005

Kane, N. C., Gill, N., King, M. G., Bowers, J. E., Berges, H., Gouzy, J., et al. (2011). Progress towards a reference genome for sunflower. Mol. Breeding 28, 1455–1462. doi: 10.1007/s11032-011-9580-1

Koike, N. C., Gill, N., King, M. G., Bowers, J. E., Berges, H., Gouzy, J., et al. (2011). Progress towards a reference genome for sunflower. Mol. Breeding 28, 1455–1462. doi: 10.1007/s11032-011-9580-1

Levi, A., and Rowland, L. J. (1997). Identifying blueberry cultivars and evaluating their genetic relationships using randomly amplified polymorphic DNA (RAPD) and simple sequence repeat (SSR) anchored primers. J. Amer. Soc. Hort. Sci. 122, 74–80.

Li, X., Sun, H., Ru, D., Dong, Y., Yang, J., Chen, H., et al. (2012). De novo sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. Genes 3, 41–64. doi: 10.3390/genes3020041

Moran, J., Neumann, P. and Navrátilová, A. (2007). Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybeans and Medicago truncatula. BMC Genomics 8:427. doi: 10.1186/1471-2164-8-427

Maden, B. (2011). EWDHE: the human encyclopedia. Nature 480, 46–48. doi: 10.1038/480048a
Rowland, L. J., Ogden, E. L., and Ehlenfeldt, M. K. (2010). EST-PCR mark-Shizuya, H., and Kouros-Mehr, H. (2001). The development and applications of
Rowland, L. J., and Levi, A. (1994). RAPD-based genetic linkage map of blueberry
Semagn, K., Bjørnstad, Å., and Ndjiondjop, M. N. (2006). Principles, requirements
Mäkinen, V., Salmela, L., and Ylinen, J. (2012). Normalized N50 assembly met-
Mudd et al. Students’ perspective on genomics
Rowland, L. J., Bell, D. J., Alkharouf, N., Bassil, N. V., Drummond, F. A., Beers, L.,
Rowland, L. J., Alkharouf, N., Darwish, O., Ogden, E. L., Polashock, J. J., Bassil, N. V.,
Osoegawa, K., Mammoseret, A. G., Wu, C., Frengen, E., Zeng, C., Catanese, J. J., et al.
Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al.
De novo genome assembly. Brief. Bioinform. 5, 257–248. doi: 10.1093/bib/bll018
Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). I renewed path approach to DNA
Pop, M., Kosack, D. S., and Salzberg, S. L. (2004a). Hierarchical scaffolding with
Pop, M., Phillippy, A., Dchel, A. L., and Altshuler, S. L. (2006b). Comparative
Rowland, L. J., Alkharouf, N., Drewil, O., Oglen, E. L., Polacheck, J. J., Basal, N. V., et al. (2012a). Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. BMC Plant Biol. 12:46. doi: 10.1186/1471-2229-12-46
Rowland, L. J., Bell, D. J., Alkharouf, N., Basal, N. V., Drummond, F. A., Beers, L., et al. (2012b). Generating genomic tools for blueberry improvements. J. Integr. Plant Biol. 12, 276–287. doi: 10.1111/j.1744-7962.2011.019452
Rowland, L. J., and Levy, A. (1994). RAPD-based genetic linkage map of blueberry
derived from a cross between diploid spcies (Vaccinium darrowi and V. elliotii).
Proc. Appl. Genet. 87, 865–868. doi: 10.1016/1045-4672(94)90098-8
Rowland, L. J., Ogden, E. L., and Hellmut, M. K. (2010). EST-PCR markers developed for highbush blueberry are also useful for genetic fingerprinting and relationship studies in rabbiteye blueberry. Sci. Hort. 125, 779–784. doi: 10.1016/j.scienta.2010.03.008
Semagn, K., Bremetad, Å., and Nilssonoy, M. N. (2006). Principles, requirements and
treating of genetic mapping in diploids. Afr. J. Biotechnol. 5, 2569–2587.
Shinya, H., Beren, B., Kim, U. J., Mancias, V., Nagak, T., Tsutoda, Y., et al. (1992).
Clossing and stable maintenancse of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89, 8794–8797. doi: 10.1073/pnas.89.20.8794
Shinya, H., and Kitos-Morde, H. (2001). The development and applications of the
bacterial artificial chromosome cloning system. KEA J. Mol. 50, 26–30. doi: 10.2520/bjm.90.26