Upregulating cytochrome P450 in *Populus trichocarpa* to Filter Out Trichloroethylene from Soil

Mercedes Goering, Ioannis Raymond, Isabella Ronchetti, Alex Zuffoletti
BioBuilder Club, Renaissance School, Charlottesville, Virginia, USA

Reviewed on 11 May 2019; Accepted on 17 June 2019; Published on 28 October 2019

Trichloroethylene or TCE (C_2HCl_3), a common industrial chemical used as a degreaser and a refrigerant, often leaks into groundwater supplies surrounding factories or developed areas. In the USA, 54% of the Environmental Protection Agency (EPA) designated Superfund sites contain TCE . According to the US Department of Health and Human Services, TCE can cause non-Hodgkin's lymphoma, kidney, and liver cancers. Research conducted in 2012 by a team led by Stan Wullschleger has suggested that planting *Populus trichocarpa*, commonly known as the poplar tree or the black cottonwood, could ameliorate this issue of groundwater pollution by TCE. Located on chromosome 1 of the tree's 19 chromosomes are cytochrome P450 genes that code for cytochrome P450 enzymes. These enzymes are known for filtering out chemical impurities in water. Additionally, the P450 enzymes of the poplar tree break down TCE into harmless byproducts such as chloride ions, a harmless salt that the tree releases, water, and carbon dioxide. However, P450 enzymes are not naturally expressed at great enough rates for poplars to make a noticeable difference in lessening groundwater pollution. This project proposes the use of clustered regularly interspaced short palindromic repeats (CRISPER) Cas9 technology, specifically, the CRISPR dCas9 (deactivated Cas9) complex, to upregulate the expression of cytochrome P450 genes, allowing poplars to remove TCE from groundwater and metabolize it into harmless byproducts at significantly increased rates compared to poplars for which the expression of cytochrome P450 genes is unaltered.

Is it possible to genetically engineer poplars in order to upregulate the expression of cytochrome P450 to make poplar trees safely remove pollutants from groundwater through phytoremediation and metabolize the pollutants into harmless byproducts?

Keywords: Poplar, cytochrome P450, trichloroethylene, upregulate, groundwater pollution, CRISPR-Cas9, phytoremediation

Mentors: Anna Minutella, Khaled Said - Authors are listed in alphabetical order.
Please direct all correspondence to: aminutella@renaissanceschool.org

Watch a video introduction by the authors at http://bit.ly/2MRp8cv

Background

In Italy, the Pontine Marshes were drained in the 1930s to make the area into farmland (Dizikes 2010). Today, this farmland has been heavily polluted by runoff and animal waste, making it difficult for people to live and cultivate agriculture (Dizikes 2010). Dr. Alan Berger from the Department of Civil and Environmental Engineering at MIT proposed in 2008 that plants, which filter out impurities and pollutants, be planted in the marsh (Dizikes 2010). After two years of campaigning, the government agreed that this was the best course of action to solve the issue of pollution in the marshes and Dr. Berger began testing and researching with his team back in Boston (Dizikes 2010). He notes in his findings from observations of the Pontine Marshes, as well as his findings in the laboratory, that trees in wetlands can decrease water pollution by up to 90% (Dizikes 2010).

Since the early 1990s, the potential of using plants to remove pollutants from groundwater by means of their roots, a process known as phytoremediation, has been clear to scientists (University of Washington 2007). The plants change those pollutants into the harmless byproducts of chloride salts, carbon, and hydrogen. Carbon and hydrogen combine with oxygen to form carbon dioxide and water, respectively. The plant uses the byproducts for itself or releases them into its surrounding environment (University of Washington 2007). Of the known plants that do this, the most prominent is the poplar tree, which is able to remove up to 3% of trichloroethylene (TCE) from the groundwater that passes by its roots (University of Washington 2007).

Only 2.5% of water on earth is fresh, and a small fraction of this water is potable (Kukreja 2017). Many countries do not have sufficient access to consumable water (Kukreja 2017). Roughly twenty percent of groundwater in China, the main source of drinking water, is heavily contaminated by industrial runoff (Kukreja 2017). Around the world, in less developed countries, 700 million people are still without access to clean drinking water, although this is an issue which has been getting better since water purification has become a major focus of humanitarian organizations (Kukreja 2017).

Groundwater pollution is an extremely pressing issue as the increase of pollutants in the environment leads to the contamination of potable water. It is increasingly necessary to find an efficient and cost-effective method to fix this problem.

There are many other options available to remove TCE from groundwater. These include more conventional techniques to remove TCE, such as excavation and transport, soil washing, addition of oxidants, incineration, extraction, pumping, and treating of contaminated water (McCutcheon and Schnoor 2003). Cometabolic aerobic and anaerobic bioremediation are also effective methods for breaking down TCE in groundwater. Cometabolic bioremediation relies on microorganisms that degrade contaminants using a cofactor that is produced while the microorganisms are reducing other compounds for energy and carbon (EPA 2006). Cometabolic bioremediation may be aerobic or anaerobic (EPA 2006). In aerobic cometabolic bioremediation, the contaminant is degraded by a cofactor produced by a microorganism during the metabolism of another compound with oxygen, whereas in anaerobic cometabolic bioremediation this process occurs without oxygen (EPA 2006). Each of these methods for removing TCE from groundwater has benefits and drawbacks, but the most important reason for why they are not heavily utilized is that the costs of establishing and managing these methods are prohibitively high. After the initial testing phases, the method of removing TCE from groundwater provided in this paper would ultimately be much more cost effective both to establish and manage.

Unfortunately, phytoremediation in its current state is too slow to be a worthwhile method for removing contaminants from highly polluted aquifers (University of Washington 2007). Poplars remove pollutants from groundwater at a much slower rate than pollutants are added to groundwater (University of Washington 2007). Additionally, poplars only remove pollutants from groundwater when they are growing, and poplar growth halts in the winter (University of Washington 2007). However, genetically modifying poplars to allow them to remove pollutants from groundwater water at much greater rates and in greater quantities, as well as allowing them to be more efficient in converting the contaminant into harmless byproducts, could provide a highly effective method of removing pollutants from groundwater (University of Washington 2007). This method would also be lower cost and lower maintenance than current methods to remove TCE from groundwater. We plan to modify the poplars using CRISPR Cas9 technology. This technology, which was discovered in bacteria and functions in a bacterial immune system, allows gene editing (Doudna and Sternberg 2017). Depending on how it is used, this technology can delete segments of a gene, insert new base pairs into a gene, or change the expression of a gene by binding to the gene and influencing the expression of the gene with outside elements that have been added to this system (Doudna and Sternberg 2017). We plan to upregulate the expression of cytochrome P450 using this last method. The technology used in this method is called CRISPR dCas9 (Doudna and Sternberg 2017). In this system, the Cas9 enzyme, which cuts DNA strands to allow for the insertion or deletion of base pair, has been deactivated (Doudna and Sternberg 2017).

This experiment focuses on the poplar tree because poplars have deep, expansive root systems which

stretch generally between 160 to 450 feet into the earth from the base of the tree, and easily reach into underground aquifers (Rockwood n.d.). Furthermore, Poplars are ideal for this experiment because they have one of the highest water uptake rates of all trees (Doty et al. 2007). The U.S. Department of Energy Joint Genome Institute (JGI) provides access to analysis tools and a public genome database. Over the course of the past few years, the JGI has sequenced over 56 trillion nucleotides of everything from microorganisms to plants (Nordberg et al. 2014). In 2014, they were able to map the genome of the poplar tree (*Populus trichocarpa* v3.0) (Nordberg et al. 2014). The poplar was sequenced ten times in order to achieve the most accurate results (Nordberg et al. 2014). It was chosen to be mapped as its sequence is significantly shorter than that of other common trees, but also, with its 19 chromosomes containing 388 mega-base pairs of sequence, the poplar genome is four times the size of the first plant to be ever mapped (Nordberg et al. 2014).

This genome was mapped thanks to Arachne version 20071016HA, a computer genome mapping system, which allowed the researchers to merge the outbred haploids on the genome, as well as to remove contaminants and integrate a clone sequence (Nordberg et al. 2014). Using Program to Assemble Spliced Alignments (PASA), transcript assemblies of many different poplar species were mapped at JGI (Nordberg et al. 2014). Using previously mapped organisms such as rice, soybeans, and fruits, JGI was able to designate the transcript assembly alignments of the genome (Nordberg et al. 2014). The gene loci were predicted using various different software, and the best prediction for each locus was selected using many positive factors and the single negative factor of overlap with repeats (Nordberg et al. 2014). The loci were then paired by aligning their proteins (Nordberg et al. 2014).

Systems Level

The goal of this project is to upregulate the expression of the cytochrome P450 genes in poplar trees in order to make poplar trees remove pollutants from groundwater and process them into harmless byproducts at greater rates. All organisms have cytochrome P450 genes which moderate and help with the synthesis and metabolism of proteins a cell needs to survive (Cytochrome P450… n.d.). It is this cytochrome P450 gene family that gives the poplar tree its ability to phytoremediate. Poplars high numbers of P450s allow them to naturally be able to remove more pollutants from groundwater than other trees (University of Washington 2007). We will be working with the black cottonwood, a poplar tree common on the west coast of North America (Poplar n.d.). CRISPR Cas9 will be used to upregulate the specific cytochrome P450 gene *CYP87A*, also called *P450 87A3*. This is the gene found in poplars that allows the tree to metabolize trichloroethylene (Wullschleger et al. n.d.).

Poplar trees produce enzymes that break down this compound into chloride ions, and recombine the hydrogen and carbon with oxygen, producing water and carbon dioxide (University of Washington 2007). The poplar tree metabolizes and degrades TCE in a pathway similar to that used in mammals (Doty et al. 2007). This system is called P450 mediated oxidative transformation (Shang et al. 2001). The precise metabolic pathway through which this metabolism occurs is not yet fully understood. At this point, scientists are only able to analyze the byproducts of P450 mediated oxidative transformation after the metabolism has occurred. Once the metabolic pathway is better understood, it will be necessary to evaluate the risks of upregulating the function of this pathway.

The promoter, which allows for gene expression, for the gene *CYP87A3* in poplars likely has some rate limiting function (University of Washington 2007). If the promoter was activated whenever it sensed specific contaminants, then the gene would always be expressed in poplars that grow above highly polluted aquifers However, that is not the case. The promoter must have some way for it to limit the expression of the cytochrome P450 enzyme associated with metabolizing TCE. It is also possible that the rate limiting function is epigenetic, or some combination of genetic and epigenetic limiters. The understanding of this function, and what the effects would be for the tree if CRISPR was used to override that function, is essential to this experiment. The next step in the experiment would be one of the following: edit the genome and override the rate-limiting function to allow for very high expressions of cytochrome P450, or discover that the rate-limiting function is in place for a very important reason and that overriding it could be dangerous to the tree's health.

In an experiment conducted by a research team at the University of Washington led by Sharon Doty that used transgenic poplar trees with increased phytoremediation function to remove TCE from goundwater, it was found that the transgenic tree grew normally, and did not display any adverse reaction to the drastic rise in the intake and metabolization of TCE (Doty et al. 2007). This suggests that it is unlikely that overriding the rate limiting function and upregulating the expression of the cytochrome P450 gene in poplars will the dangerous to the trees' health. However, there is a possibility that the use of CRISPR dCas9 to override the rate limiting function could result in off-target effects deleterious to the trees' health since dCas9 deactivates the gene to which it bonds and essentially becomes an epigenetic tag on that gene. If it is decided that it is safe to override the rate limiting function of the promoter, and that no harm-

ful off-target effects will result from using CRISPR to do this, then we would have to discover what part of the promoter causes that rate limiting function, and how that segment could be edited using CRISPR to override the rate-limiting function.

Device Level

This experiment looks to upregulate expression of the poplar gene cytochrome *P450 87A3*. This gene controls cytochrome P450 and related proteins that are used in the phytoremediation process, particularly with regard to the metabolization of TCE. *P450 87A3* is located on Chromosome 1 of 19 (Cytochrome P450 n.d.; The Populus Genome…, n.d.). To upregulate this gene, we will use a specific type of CRISPR system called CRISPR dCas9. dCas9, or deactivated Cas9, is a Cas9 protein that does not cut out sections of DNA, but can still be programmed to fuse to the DNA and attract or repel transcription initiation factors in order to upregulate or restrict expression (Gilbert et al. 2013). By fusing the dCas9 protein to the effector domains of cytochrome P450 genes we can upregulate the expression in a controllable and stable fashion (Figure 1; Gilbert et al. 2013). The dCas9 protein can be used to very specifically target effector domains on the DNA of the poplar and not only repress but also allow upregulation of DNA binding and expression (Gilbert et al. 2013).

It has been shown that dCas9 can be efficiently used to regulate the expression of proteins (Gilbert et al. 2013). dCas9 can be sent to a promoter which couples with CRISPR guide RNA, and it can either block the promoter, or attract other proteins to upregulate the promoter (Gilbert et al. 2013). Another way in which Cas9 can be used to regulate genetic output is by means of developing catalytically inactive Cas9 (Gilbert et al. 2013). In a study by Le Sage and Cross (n.d.) called "CRISPRa: Transcriptional upregulation screening with genome-wide

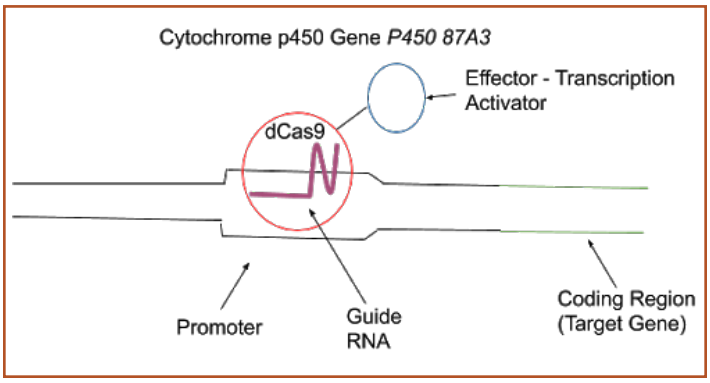

Figure 1. *The dCas9 system uses a transcription activator to induce upregulation*

CRISPR activation," researchers used CRISPRa, *activation*, in order to upregulate the production of a certain resistance gene. While this gene is virtually irrelevant to our poplar study, the technique used in this Cambridge University study is.

The first step necessary in completing this experiment is locating the gene *P450 87A3* on the poplar genome. Next, the CRISPR dCas9 system will be inserted into the poplar genome so as to target the promoter of the cytochrome P450 gene and upregulate it. The system provides elements of stability, as the targeting of genes is very precise, and the gene is activated by a variety of elements across the gene involved in regulation of expression.

Parts Level

The 19-chromosome long genome of P. trichocarpa v3.0 was sequenced in 2003 and released to the public in 2006 (Wullschleger et al. 2012). The CYP87A3 gene, which produces the cytochrome protein that metabolizes trichloroethylene, is located on chromosome one (Wullschleger et al. 2012).

This experiment looks to use the CRISPR dCas9 system to upregulate the promoter of CYP87A3. CRISPRs are the most widely shared family of repeating DNA sequences in all prokaryotes, with about half of the bacterial genomes that have been sequenced to date including CRISPRs, and almost every archaeal genome sequenced to date including CRISPRs (Doudna and Sternberg 2017). CRISPRs are exact copies of the genetic information of various bacteriophages, or viruses that infect bacteria (Doudna and Sternberg 2017). The more CRISPR spacers a bacteria has, the less likely that bacteria is to get infected by a virus (Doudna and Sternberg 2017). The CRISPR spacers, along with the CRISPRs themselves, are part of a bacterial immune system (Doudna and Sternberg 2017).

CRISPR-associated genes, or Cas genes, are a set of genes that almost always flank the CRISPR regions of bacterial genomes (Doudna and Sternberg 2017). Cas genes code for specialized enzymes, which function to unzip the DNA double helix or cleave RNA or DNA molecules (Doudna and Sternberg 2017).

The Cascade molecular machinery, which functions as a bacterial immune system, contains CRISPR RNA and 10 or 11 different Cas proteins (Doudna and Sternberg 2017). The Cascade molecular machinery works by latching on to its viral DNA targets and determining the exact sequence of viral DNA that is to be destroyed (Doudna and Sternberg 2017). Cascade is very adept at locking onto perfect or near perfect matches of the CRISPR RNA,

and its high level of discrimination ensures that it does not accidently target a sequence in the bacterial DNA that is similar to the CRISPR RNA sequence (Doudna and Sternberg 2017).

In the CRISPR-Cas9 system, CRISPR RNA plays the role of recognizing a sequence of viral DNA that is identical to the sequence of the CRISPR RNA (Doudna and Sternberg 2017). CRISPR RNA also plays the role of bonding to one of the strands of viral DNA, forming a DNA-RNA double helix (Doudna and Sternberg 2017). TracrRNA performs the simple but necessary task of holding the CRISPR RNA - Cas9 complex together (Doudna and Sternberg 2017). Cas9 plays the role of separating the viral DNA double helix, allowing for the CRISPR RNA to form an RNA-DNA double helix with one of the viral DNA strands (Doudna and Sternberg 2017). Cas9 then performs the task of snipping both of the strands of viral DNA using two nuclease modules, initiating a double-strand break (Doudna and Sternberg 2017).

Cas9 (160 kDa) is an enzyme, meaning that it is built up of hundreds to thousands of amino acids (Doudna and Sternberg 2017). However, nearly all of these amino acids are there only to define the enzyme's shape. Only a few contribute to the actual biochemical function of the enzyme (Doudna and Sternberg 2017). Deactivating those few amino acids in Cas9 that allow it to snip DNA takes away the DNA cutting ability of Cas9, but leaves Cas9 still able to interact with the RNA guiding sequence and attach itself to a matching DNA sequence (Doudna and Sternberg 2017). By adding proteins that influence the output of genes to the CRISPR-Cas9 complex with a deactivated Cas9 enzyme, scientists can control gene expression (Doudna and Sternberg 2017).

We do not have enough information about the promoter of the *P. trichocarpa v3.0* gene to go into further detail at this point in the discussion of the parts level. The next steps would be to thoroughly examine the promoter of the *CYP87A3* gene, and to use the information from this analysis to determine which proteins should be added to the CRISPR dCas9 complex in order to upregulate the expression of *CYP87A3.* It would then be necessary to investigate the method by which these proteins are added to the CRISPR dCas9 complex, as well as the method by which this complex is most effectively inserted into a poplar cell to cause the upregulation of *CYP87A3* to occur.

Safety

This experiment has minimal safety risks, but how it may be perceived by the public is a question which requires further investigation. Federal regulations allow genetically modified trees to be grown for research purposes in greenhouses or controlled fields, but they are not permitted to be grown for commercial purposes (University of Washington 2007). The one major safety risk regarding genetically engineered trees is how they might affect the ecosystem that they enter (University of Washington 2007). How might bugs and other animals that feed on the leaves and twigs of trees be affected by a genetically engineered poplar grove? What if some of the trees experience off-targeting effects from being genetically modified that cause harm to not just the tree, but the surrounding environment. The main worry surrounding these trees is what may happen if seeds from the genetically engineered poplars spread before it can be discovered if they have any negative effects on their surrounding environment (University of Washington 2007). The ultimate goal is to remove carcinogens and other harmful substances from water supplies, so it makes sense to do as much testing as possible to ensure that nothing is being introduced into the environment that would be even worse than carcinogens in drinking water (University of Washington 2007). If the trees would be better off not being allowed to spread, there is still potential for them to be used. Poplars grow quickly, but grow for several years before flowering (University of Washington 2007). It could be possible to plant genetically engineered poplars above highly polluted aquifers, let the poplars grow and remove many contaminants from the aquifer, and then remove the poplars before they flower and spread (University of Washington 2007). Even if someday the trees are approved and permitted to be used for phytoremediation, they will undoubtedly require close monitoring, even as nothing more than a simple precaution (University of Washington 2007).

Discussion

While there are worries about how the general public will receive the genetic modification of trees, there are no other foreseeable issues with gene expression when the promoter of the CYP87A3 gene is upregulated (University of Washington 2007). This project examines the upregulation of the production of the cytochrome P450 gene CYP87A3 in poplar trees using the CRISPR dCas9 system. This will allow the trees to break down TCE taken in from groundwater at much greater rates than poplars are naturally able. The poplar trees break down the TCE into chloride ions, carbon, and hydrogen, and the carbon and hydrogen are then combined with oxygen to form CO_2 and water. Some of these harmless byproducts the tree uses for itself, while others can even be beneficial to the surrounding environment, (University of Washington 2007). Hopefully, the general public will understand that this specific genetic modification using the CRISPR dCas9 system to upregulate expression of the CYP87A3 gene is safe and aims to clean up the freshwater of our world.

Acknowledgements

We would like to thank our Synthetic Biology teacher, Mrs. Anna Minutella, and our graduate school mentor, Khaled Said.

This project was accomplished through participation in the BioBuilderClub, an after-school program organized by BioBuilder Educational Foundation. BioBuilderClub engages high school teams around the world to combine engineering approaches and scientific know-how to design/build/test their own project ideas using synthetic biology.

References

Acosta L. Restrictions on Genetically Modified Organisms: United States. [Internet] Independence Ave (WA): The Library of Congress; 2014 Mar 1 [cited 2019 March 3] Available from http://bit.ly/2YSn0Dr

Dizikes P. Using plants to purify canal water. [Internet] Cambridge (MA) - Massachusetts Institute of Technology. 2010 Apr 7 [cited 2019 Jan 24] Available from http://bit.ly/2YEbT5S

Doty SL, James CA, Moore, AL, Vajzovic A, Singleton GL, Ma C, Khan Z, Xin G, Kang JW, Park JY, Meilan R, Strauss SH, Wilkerson J, Farin F, Strand SE. Enhanced phytoremediation of volatile environmental pollutants with transgenic trees. Proceedings of the National Academy of Sciences. 2007 Oct 23; 104 (43): 16816-21. DOI:10.1073/pnas.0703276104

Doudna J, Sternberg SH. A crack in creation. Boston, MA: Houghton Mifflin Harcourt. 2017 Jun 13.

Environmental Protection Agency. Engineering Forum Issue Paper: In Situ Treatment Technologies for Contaminated Soil. 2006 Nov. EPA 542/F-06/013.

Genetics Home Reference. Cytochrome P450 - Genetics Home Reference - NIH. [Internet] Bethesda (MD): US National Library of Medicine; (n.d.) [cited 2019 March 17]. Available from http://bit.ly/2ZaUq0r

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, . . . Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. [Internet] 2013 Jul 18 [cited 2019 Feb 19] Available from http://bit.ly/2ON875K

Gonthier BP, & Barret LG. In-vitro spin-trapping of free radical produced during trichloroethylene and diethyl ether metabolism. Toxicology letters. [Internet] 1989 Jun [cited 2019 May 23] Available from: http://bit.ly/33ks8nj

Kukreja R. 40 interesting facts about water pollution. [Internet] Conserve Energy Future; 2017 [cited 2019 Feb 26] Available from http://bit.ly/2KDQnnY

Le Sage C, Cross BCS. CRISPRa: Transcriptional upregulation screening with genome-wide CRISPR activation. Cambridge (UK): Horizon Inspired Cell Solutions. (n.d.)

McCutcheon SC, Schnoor JL. Phytoremediation: transformation and control of contaminants. Hoboken (NJ): John Wiley & Sons, Inc. 2003.

National Center for Biotechnology Information. Cytochrome P450 87A3 [Populus trichocarpa] - Protein - NCBI. [Internet] Bethesda (MD): National Center for Biotechnology Information; (n.d.) [cited 2019 April 15] Available from: http://bit.ly/2MP1du4

Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. [Internet] Walnut Creek (CA): Joint Genome Institute; 2014 [cited 2019 May 2] Available from http://bit.ly/33n6dvR

Rockwood P. (n.d.). How Far Do Poplar Roots Spread? [Internet] San Francisco (CA): SFGate; (n.d.) [cited 2019 Feb 28] Available from: http://bit.ly/2KDQJuO

Shang, TQ, Doty SL, Wilson, AM, Howald, WN, Gordon, MP. Trichloroethylene oxidative metabolism in plants: the trichloroethanol pathway. Phytochemistry. 2001 Dec. 58(7): 1055-65 University of Washington. Genetically engineered poplar plants disarm toxic pollutants 100 times better than controls. [Internet] Rockville (MD): ScienceDaily; 2007 Oct 16 [cited 2019 April 5] Available from http://bit.ly/2YAxMmP

Woodland Trust. Poplar, black (Populus nigra). [Internet] Lincolnshire (ENG): Woodland Trust; (n.d.) [cited 2019 Feb 28] Available from http://bit.ly/2Z8vdI1

Wullschleger SD, Weston DJ, DiFazio SP, Tuska, GA. Revisiting the sequencing of the first tree genome: Populus trichocarpa. Tree physiology. [Internet] 2012 Oct 25 [cited 2019 March 9] Available from: http://bit.ly/2Kp0PAL