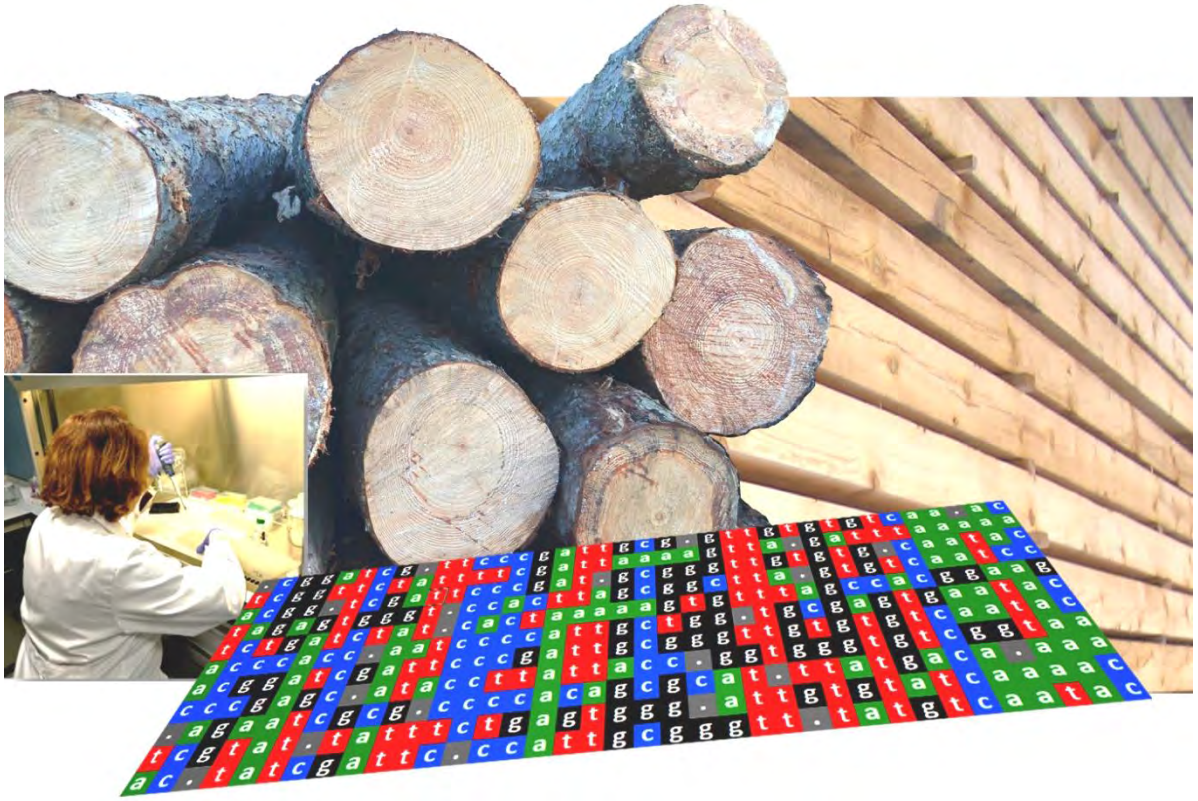
OECD
BETTER POLICIES FOR BETTER LIVES

INIA
Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria

Universidad de Alcalá

GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA Y PESCA, ALIMENTACIÓN Y MEDIO AMBIENTE

Application of high-throughput genotyping technologies for forest tree species identification and timber tracking



Editors: M.T Cervera, J.A. Cabezas, C. Díaz-Sala

Application of high-throughput genotyping technologies for forest tree species identification and timber tracking

Edited by

Maria Teresa Cervera

Centro de Investigación Forestal (CIFOR), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain.

José Antonio Cabezas

Centro de Investigación Forestal (CIFOR), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain.

Carmen Díaz-Sala

Universidad de Alcalá, Alcalá de Henares, Spain.

Published by

Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente

The Workshop was sponsored by the OECD Co-operative Research Programme on Biological Resource Management for Sustainable Agricultural Systems, whose financial support made it possible for most of the invited speakers to participate in the Workshop.

Other sponsors:

ACKNOWLEDGEMENTS

We would especially like to thank the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems for supporting and funding the organization of the workshop *"Application of high-throughput genotyping technologies for forest tree species identification and timber tracking"*. We would also like to express our gratitude to Instituto Nacional de Investigaciones Agrarias y Alimientarias (INIA) and Universidad de Alcalá (UAH UE2017-002) for co-funding the organization of the workshop. Ministerio de Agricultura, Pesca. Alimentación y Medioambiente (MAPAMA) provided the facilities for the organization of the workshop and funded this book.

PREFACE

Technical advances in genome sequencing and genotyping over the past 20 years, have contributed to document the genetic variability of a large set of plant species, including those with large and complex genomes such as different forest trees. The combination of next generation sequencing with computational applications have allowed selecting minimal sets of markers per species, avoiding interference, with which to design multispecies genotyping tools with high discrimination capacity. Such tools may help to increase the tracking efficiency of timber species threatened by illegal logging and trading. This is a key application to prevent illegal logging and associated trade.

This book summarizes the topics discussed in the workshop *"Application of high-throughput genotyping technologies for forest tree species identification and timber tracking"*, which was held in Madrid, September 13-15, 2017, sponsored by the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems. The workshop aimed at gathering international stakeholders involved in forest tree species identification and timber tracking. Organized at the initiative of researchers involved in international forest tree genomic initiatives, the programme provided room for extensive discussion on appropriate strategies to efficiently transfer information and experience to standardize sampling and to identify informative markers from timber species threatened by illegal logging for which genotyping tools lacked. The structure of databases and data availability was also amongst key topics addressed.

The structure of this book follows the structure of the sessions and presentations included in the workshop programme. Presentations provided detailed updated information about legal, economic and environmental issues. Scientific and technological advances about high-throughput technologies to identify tree variability were provided, including status of publically available and sharable data. The bioinformatic tools needed to manage data and models to build a common database were also addressed. Available case studies were analyzed to identify and prioritize hot-spots to be addressed in the future.

M.T. Cervera, J.A. Cabezas, C. Díaz-Sala

Madrid, January 2018

CONTENTS

- **PART 1. ECONOMICAL, LEGAL AND ENVIRONMENTAL ASPECTS OF THE TIMBER ILLEGAL LOGGING AND TRADING** **15**
  - Chapter 1.1. *Shelley Gardner*
    US Lacey Act 2008 Amendments 17
  - Chapter 1.2. *Claudine Leger-Charnay*
    INTERPOL's Project LEAF (Law Enforcement Assistance for Forests) against forestry crime: strategy and activities 23
  - Chapter 1.3. *Steven Johnson*
    International cooperation on tropical forests: Governance, legality and ITTO 29
  - Chapter 1.4. *Jo Van Brusselen*
    The Global Timber Tracking Network 35
  - Conclusions 41
- **PART 2. HIGH-THROUGHPUT TECHNOLOGIES TO IDENTIFY TREE VARIABILITY** **43**
  - Chapter 2.1. *Antoine Kremer*
    A set of genomic and electronic resources to discriminate genetic units in European white oaks 45
  - Chapter 2.2. *Nathalie Isabel, et al.*
    Towards the development of a traceability system for Canadian forest products 49
  - Chapter 2.3. *Valerie D. Hipkins*
    Use of genotyping technologies to identify tree variability 55
  - Chapter 2.4. *Patricia Favre-Rampant et al.*
    How New Generation Sequencing and genotyping technologies can help development of DNA traceability tools 59
  - Conclusions 63
- **PART 3. BIOINFORMATICS ANALYSIS AND DATABASE GENERATION BASED ON HIGH-THROUGHPUT INFORMATION** **51**
  - Chapter 3.1. *José De Vega*
    Cost-effective approaches for variant calling and analysis of complex plants 65
  - Chapter 3.2. *Jill L. Wegrzyn*
    Timber Tracking Initiatives -Computational Support with TreeGenes 75
  - Chapter 3.3. *Lieven Steck*
    Bioinformatics tools and resources to study plant genomes 79
  - Chapter 3.4. *Tommi Suominen*
    Developing an expert database and a reference database with the Global Timber Tracking Network 85
  - Conclusions 91
- **PART 4. APPLICATION OF DNA TECHNOLOGIES TO PREVENT TIMBER ILLEGAL LOGGING AND TRADING** **93**
  - Chapter 4.1. *Bernd Degen and Céline Blanc-Jolivet*
    Development and application of genetic reference data based on SNPs for timber tracking of tropical tree species 95
  - Chapter 4.2. *Andrew J Lowe et al.*
    Opportunities for improved transparency in the timber trade through advanced DNA analysis 101
  - Chapter 4.3. *Stephen Cavers*
    Challenges to implementation of high-throughput genotyping technologies for DNA forensics in the timber market 109
  - Chapter 4.4. *Marius R. M. Ekué et al.*
    SNPs based timber tracking tools for African mahogany *Khaya* sp. 115
  - Conclusions 117
- **FINAL OUTCOMES. SUMMARY AND PROPOSED ROADMAP** **119**

PART 1

ECONOMICAL, LEGAL AND ENVIRONMENTAL ASPECTS OF THE TIMBER ILLEGAL LOGGING AND TRADING

Chapter 1.1.

US Lacey Act 2008 Amendments

Shelley Gardner
U.S. Forest Service, International Programs, USA

The Lacey Act is a 1900 United States law that bans trafficking in illegal wildlife. In 2008, the Act was amended to include plants and plant products such as timber and paper. This landmark legislation is the world's first ban on trade in illegally sourced wood products.

There are two major components to the plant amendments: a ban on trading plants or plant products harvested in violation of the law; and a requirement to declare the scientific name, value, quantity, and country of harvest origin for some products.

The Lacey Act is a fact-based statute with strict liability, which means that only actual legality counts (no third-party certification or verification schemes can be used to "prove" legality under the Act) and that violators of the law can face criminal and civil sanctions even if they did not know that they were dealing with an illegally harvested product.

Penalties for violating the Lacey Act vary in severity based on the violator's level of knowledge about the product: penalties are higher for those who knew they were trading in illegally harvested materials. For those who did not know, penalties vary based on whether the individual or company in question did everything possible to determine that the product was legal. In the U.S. system, this is called "due care," and is a legal concept designed to encourage flexibility in the marketplace.

More information from various U.S. government sources regarding the definition and exercise of due care include the following "Lacey Act Primer", a presentation from the USDA Animal and Plant Health Inspection Service (slides 17-21)

Enforcement

The Lacey Act has a long history of successful enforcement as a wildlife statute, and over a century of case law on these older provisions of the Act is readily available. Three examples of enforcement cases that have used the plant product amendments: two regarding a major U.S. guitar manufacturer, and one related to a small business.

Other U.S. Government Documents and Links

ThePlant and Plant Product Declaration Form is required for many types of plant and plant product imports into the United States.

U.S. Customs and Border Protection has information on the amended Lacey Act, along with useful CBP guidance on Lacey Act declarations.

The Animal Plant Health Inspection Service (APHIS) of the U.S. Department of Agriculture is a primary implementing agency for the amended Lacey Act. See APHIS's website for a wealth of information on the Lacey Act, including FAQs, guidance on import declarations, and direct contact information. The site also offers the opportunity to be registered as a stakeholder in the declaration requirement implementation process and receive regular updates from APHIS.

Best Practice Guide for Forensic Timber Identification

In order to ensure that forensic data are credible and admissible in court, appropriate methods and procedures must be used throughout the entire investigative process, from the first inspection of a timber load, to timber sample collection and transport, analysis in the laboratory, and interpretation and presentation of results for prosecution.

This Guide is intended for worldwide use, with the aim of facilitating the employment of forensic science to the fullest extent possible to combat timber crime. This Guide covers the whole chain of events, providing information on best practices and procedures from the crime scene to the court room. The target audience ranges from front-line officers, crime scene investigators, law enforcement officials, scientists, prosecutors and the judiciary. The Guide, as a whole, represents a starting point for a uniform approach to the collection and forensic analysis of timber for identification purposes. It is hoped that the use of the Guide will lead to more timely, thorough and effective investigations, resulting in an increased number of successful prosecution and a reduction in the illegal timber trade.

Due to the varied, complex and highly technical nature of timber identification methodologies, this Guide does not provide step-by-step scientific processes for their application in the field or laboratory. Instead, this Guide focuses on the procedural aspects for obtaining robust identification outcomes suitable for presentation in court to support illegal timber trading prosecutions. A glossary of terms can be found in annex 1. Example resources detailing the required scientific methodologies are referred to throughout the Guide; however, as forensic timber identification is a growing discipline, resources cited here should be considered only as examples. To obtain a current picture of the available resources, a forensic timber identification expert should be consulted.

Wood can be processed in a myriad of different ways; it can be turned into pulp to make paper, powdered for traditional medicine, planed into extremely thin veneers, fixed together to make plywood or worked into high value objects such as musical instruments. The applicability of the various available timber identification methodologies can vary according to the wood material in question. To avoid confusion, this Guide focuses on the identification of solid timber only. An explanation of the various other wood products that may be encountered and considerations for obtaining forensic identification for these materials can be found in annex 2. Specific information about non-solid-timber wood products of CITES-listed tree species can be found in annex 3.

The provision of forensic services is affected by the legal framework in place and includes issues related to entering the crime scene, conducting the investigation, handling evidence, laboratory analysis and others.

The Guide is divided into four parts containing information specific to different audiences. They are collectively intended to provide integrated tools for gathering and processing evidence on timber crime and performing laboratory analysis in support of prosecution and for intelligence purposes. A full reading of the Guide will provide valuable insight and advance understanding of the forensic challenges facing each actor along the crime chain.

- Part I provides information for law enforcement. It describes initial risk analysis and search guidelines for front-line officers. It advises on options for rapid-field identification and formulation of forensic questions. Guidance is provided on the collection and preservation of evidence, maintaining the chain of custody, including through transport of samples to the laboratory. It also advises on communication with the timber identification service provider.
- Part II is aimed at scientists undertaking forensic identification tests or those who seek to do so in the future. Some information is also relevant to research scientists involved in the development of identification methodologies but who may not necessarily undertake forensic case work. The various methods of timber identification are summarized as an introduction to the associated disciplines. Resources for acquiring reference material and data are presented, and guidance is provided regarding laboratory procedural requirements for undertaking forensic work. It also advises on communication with law enforcement and communication of scientific results by an expert witness in court.
- Part III is aimed at law enforcement, prosecutors and the judiciary. It is focused on appropriate considerations when preparing an illegal timber case for court. To facilitate understanding of identification methods and results by the prosecution and judiciary, simple descriptions of the relevant methods are provided. Key forensic requirements and specific legal considerations regarding the use of forensic timber identification services are discussed, and a final checklist is presented.
- Part IV discusses the importance of international cooperation to tackle timber crime. It covers relevant international legal frameworks, which form the basis for cooperation between countries, and at the global level, the basis for regulation, communication, exchange of information and mutual assistance to tackle transnational organized crime. Information is provided on networks, mechanisms and tools available for countries and individuals seeking to obtain legal or scientific assistance from another country. It outlines some of the benefits, challenges and opportunities to improve cooperation, communication and collaboration internationally among and between legal and scientific communities.

Accompanying the Guide, a best practice flow diagram (as shown in Figure 1) has been developed to lead front-line officers through the steps that should be completed when dealing with a load or shipment containing timber that is passing through a checkpoint such as an international border crossing. An online version of this flow diagram that includes dynamic links to additional resources can be accessed at www.unodc.org/documents/Wildlife/Timber_Flow_Diagram.pdf.

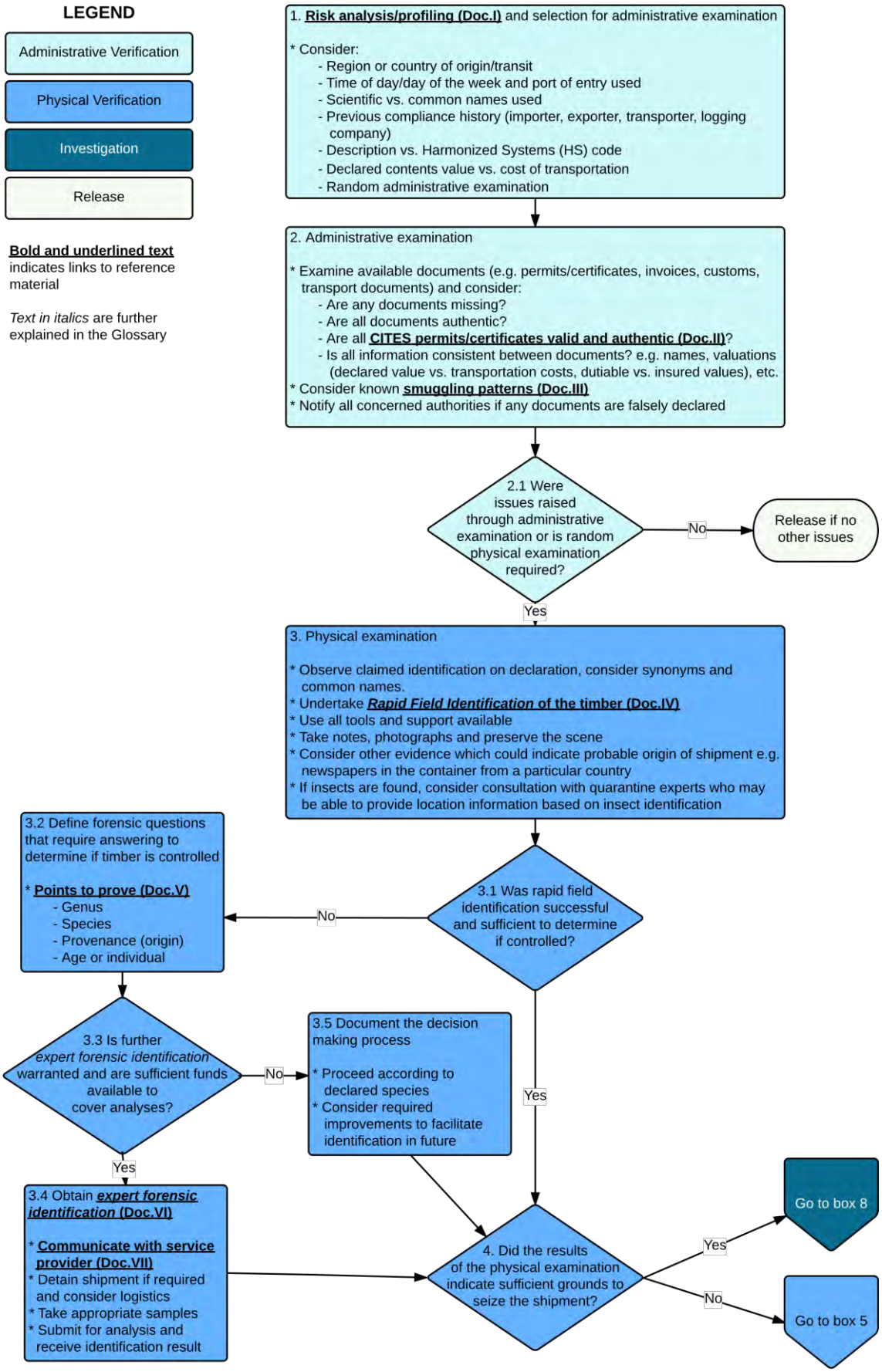


Figure 1. *Best practice flow diagram*

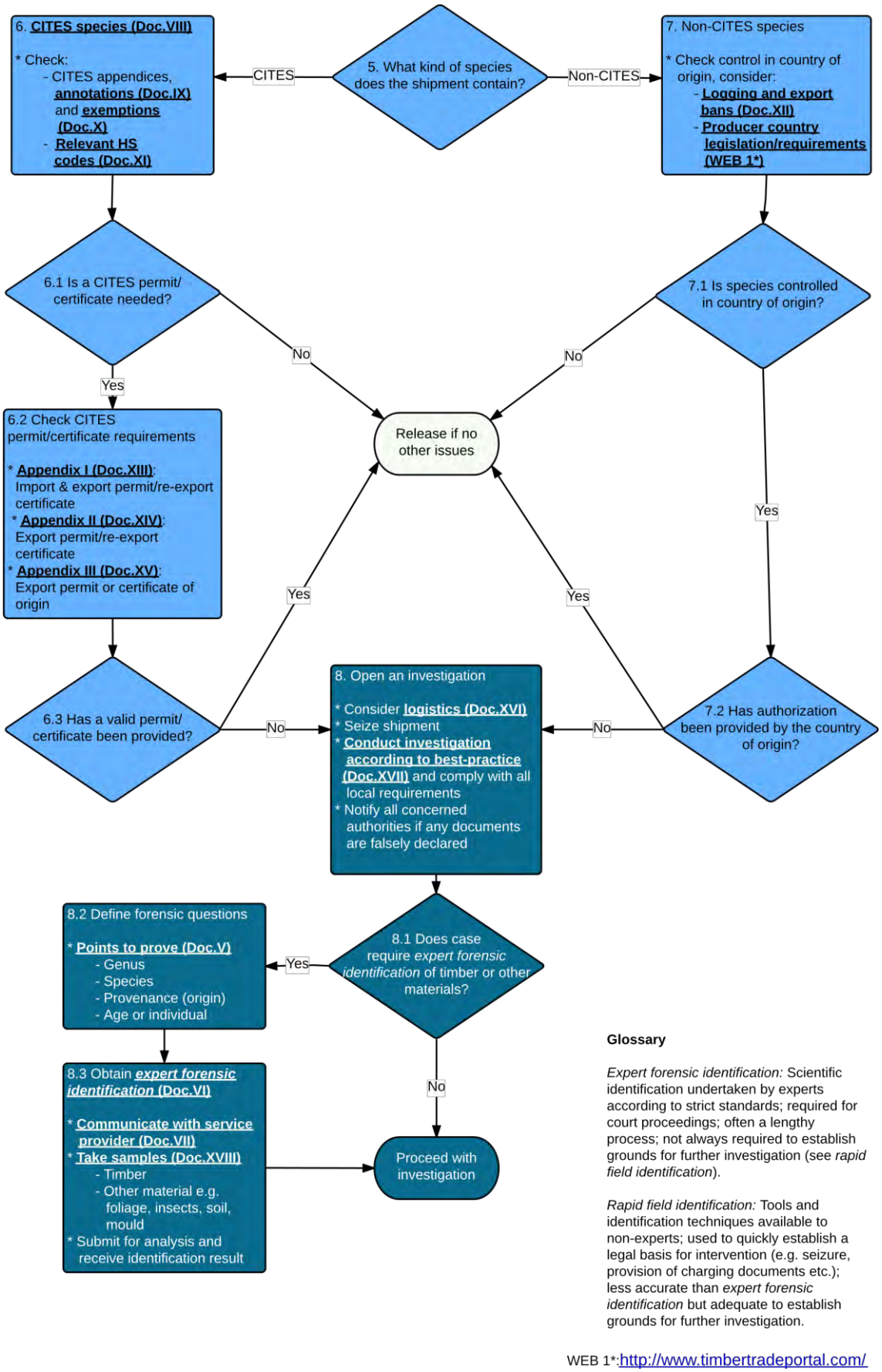


Figure 1 *(continuation)*

Chapter 1.2.

INTERPOL's Project LEAF (Law Enforcement Assistance for Forests) against forestry crime: strategy and activities

Claudine Leger-Charnay
INTERPOL, Environmental Security Programme, France

Background

- Project LEAF (Law Enforcement Assistance for Forests) is a global project: since the creation of the project in 2012 on World Environment Day, we have worked with the main timber exporting countries in Central and South America, West Africa and South-East Asia. We have also worked with the major timber importing countries, i.e. the US, the EU and Asia.
- To achieve the project's goals INTERPOL is working closely with the United Nations Environment Programme (UNEP) and the International Consortium for Combatting Wildlife Crime (ICCWC) comprising INTERPOL, UNODC, the CITES Secretariat, World Customs Organisation and the World Bank.
- Funding stream from NORAD (Norwegian Agency for Development Cooperation) and additional support from US Department of State for trainings and to support investigations.

Project Aim

- Identify and dismantle criminal networks involved in illegal logging.
- Focus on high-level criminals and heads of criminal networks.

Project Strategy

Our strategy is represented by this Project cycle shown on Figure 1:

- Illegal logging has not traditionally been high priority amongst law enforcement officials who have focused on areas such as drug trafficking or firearms.

- By raising awareness about illegal logging, Project LEAF encourages member countries and CSOs to share criminal intelligence with INTERPOL. We support member countries in analyzing criminal intelligence to identify criminals, the businesses they are linked to and their modus operandi. To this end, we organize **Regional Investigative and Analytical Meetings** or **RIACMs**. They allow investigators from member countries to meet face-to-face to review case files and share intelligence and analysis to further their investigations.
- We identify countries' capacity needs and train law enforcement officials to improve their ability to undertake targeted law enforcement operations.
- We organize transnational operations targeting Latin America, Africa and the Asia-Pacific region and provide investigative support. More specifically, INTERPOL can support the deployment of an **Investigative Support Team** (IST). An IST can be deployed at the request of a member country. This is a team of specialized law enforcement experts on digital forensics, species identification techniques, language and technical support in interviewing suspects, database queries, criminal intelligence analysis, etc. and can provide advice to issue **INTERPOL Notices** (which are alerts to member countries to share critical crime-related information);
- We disseminate recommendations and best practices for combating forestry crimes.
- All activities collectively help to strengthen national and regional law enforcement networks.

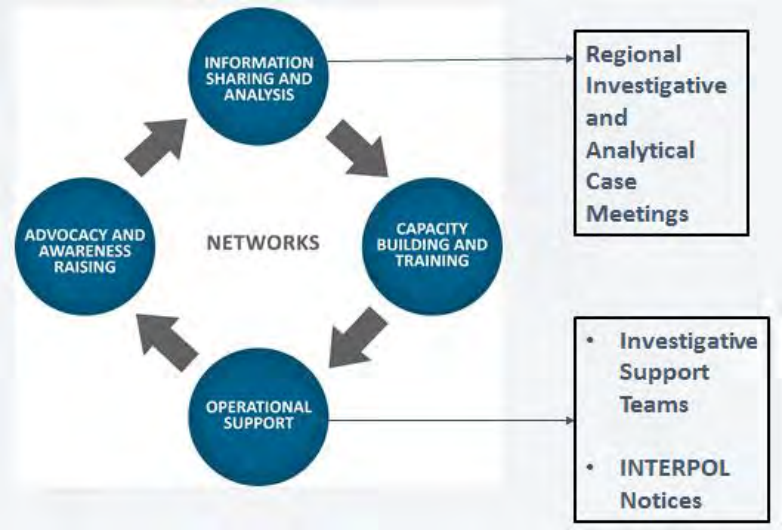


Figure 1.

NESTs (National Environmental Security Taskforces)

- Criminals exploit the lack of communication between law enforcement agencies both within and between countries. Forestry crime is organized crime so it can only be countered by multi-agency international cooperation and full supply chain traceability.

- A NEST is a group of law enforcement agencies including the police, customs, prosecutors as well as environmental agencies, IGOs and NGOs to bring a multi-disciplinary and coordinated answer to environmental crime. NESTs benefit from utilizing INTERPOL National Central Bureaus to access INTERPOL's tools and services.
- In Latin America, we have implemented 8 NESTs to date.

Project LEAF Achievements in 5 years

- 11 law enforcement operations
- 34 participating countries
- 800 law enforcement officers trained
- Volume of timber seized is equal to 1.5 bn USD, 950,000 truckloads, 570 Olympic sized swimming pools
- More than 547 arrests

Operation Amazonas II (latest operation in Latin America)

- Dates: 2015- 2016
- Objectives: Identify and dismantle international criminal networks involved in illegal trade of timber; continue transnational investigations of past and current cases related to illegal logging and illegal timber trade using INTERPOL policing capabilities, notably notices and diffusions; identify and monitor the main illegal timber transport routes and its trading hubs with a view to inspect the truck's load, shipments and containers.
- Operation led by the law enforcement agencies (INTERPOL National Central Bureaus, national police specialized in environmental crime, Customs, Prosecutors) and environmental and governmental agencies from 12 Central and South America countries including Argentina, Bolivia, Brazil, Colombia, Costa Rica, , Ecuador, El Salvador, Guatemala, Honduras, Paraguay, Peru and the Dominican Republic.
- INTERPOL brings operational, investigative and analytical support to its member countries. Results of operations are always analysed and used to build the next operation.
- Some findings of our analysis are as follow. In total, out of 131 timber species identified by Operation Amazonas II countries, only 5 are protected by CITES, including: Black rosewood, Cedar, Big- Leaf Mahogany, Caribbean Mahogany, Granadillo rosewood. This analysis suggests law enforcement should focus their efforts not only on CITES protected species. We also did analysis on timber trade routes or cross-over crimes. Countries identified document fraud as facilitating all stages of the timber supply chain from harvest, transport, processing and export to sale. Corruption was suspected to most commonly facilitate the harvest and export of timber. Illegal wildlife trafficking was also identified as a cross-over crime. Therefore, the law enforcement agencies responsible for preventing, investigating and prosecuting these different crimes are encouraged to work together to bring a multi-agency and coordinated response. Countries are encouraged to create NESTs.

Issues and solutions around timber tracking

- A robust timber tracking system is important to prevent illegal timber from entering the supply chain. Law enforcement most frequently use non-inherent features of wood to track timber: paper based certificates, painted markings, plastic tags or barcodes. Issues: susceptible to forgery (mislabelling), misinterpretations, not durable (can be detached from logs). Solution: DNA timber identification could be used in official investigations to identify species, their origin, their age, the trafficking routes (country of origin, transit and destination country), the modus operandi and audit timber tracking systems using barcodes and tags.
- An example of a case study was given to open-up for questions from the law enforcement community to the scientific community. These questions included how much time DNA analysis takes as it is costly to detain timber, how quickly can Science identify that a timber specie declared on a certificate is not this timber specie, how much money it costs as developing countries have little financial resources, how precise it is (genus or species level? Country, region, forest, concession level?), how reliable it is and if there is a unique international database (Who hosts? Who has access to? Risk of being accessed by criminals). Law enforcement need to be involved in the debates to better understand available services.

Many of these questions have been answered by the UNODC Best Pratice Guide For Forensic Timber Identification that Shelley Gardner mentioned in her presentation. The Timber Guide explores options for further development of forensic best practices to provide evidence-based information, support law enforcement investigations and lead to successful prosecutions.

INTERPOL's role to advance DNA timber identification

- INTERPOL can be a plateform to host trainings or could support training activities in member countries to develop and apply timber identification methods to support countries' investigations.
- INTERPOL could encourage member countries to take samples from seized or detained timber.
- INTERPOL encourages the participants' involvement in the INTERPOL Forestry Crime Working Group. It is a global strategic advisory body that will provide strategic advice to INTERPOL, in order to improve the effectiveness of law enforcement operations targeting organized criminal networks engaged in illegal logging and international trade in illegal timber and related crimes.

Context

Workshop on Application of high-throughput genotyping technologies for forest tree species identification and timber tracking, Madrid, Spain, 13-15 September 2017; Panel on "Economical, legal and environmental aspects of the timber illegal logging and trading". In this

panel, discussions reflected on law enforcement concerns in the development of DNA timber identification (how much time and money does DNA analysis cost, how precise and reliable are the results of the analysis, creation of an international database, etc.). INTERPOL's role in the development of DNA timber identification tools for law enforcement was also explored.

It is useful INTERPOL is represented at this kind of meetings as it allows us to better understand available services and give direction to the scientific community on how to develop forensic tools for timber identification useful for the law enforcement community.

INTERPOL will be able to raise awareness on these technologies with its member countries but local law enforcement should be invited too at future similar meetings.

Challenges on the development and use of DNA tools to identify timber species and their origin were discussed and many questions are still to be solved. However, the main aim of this meeting has been achieved which was to put in contact the scientific and governmental/law enforcement actors to explore opportunities of cooperation on these subjects.

Follow-up actions by INTERPOL

- Follow more closely the work of the Global Timber Tracking Network (GTTN).
- Continue exploring INTERPOL's role in the development and application of DNA timber identification tools jointly with GTTN.
- Invite relevant representatives of GTTN to attend some open sessions of the soon-to-be-established Forestry Crime Working Group. Since national law enforcement representatives will be the main stakeholders represented in the working group, this should contribute to raise their awareness about timber forensic identification.

Chapter 1.3.

International cooperation on tropical forests: Governance, legality and ITTO

Steven Johnson
International Tropical Timber Organization, Japan

1. ITTO and forest governance

ITTO is an international organization with 73 member countries accounting for all major global producers and consumers of tropical timber. The organization's mandate is to promote sustainable forest management (SFM) in the tropics and trade in sustainably produced tropical forest products. Recognizing that a key requirement of sustainability is compliance with all relevant legal frameworks, ITTO began work on forest governance and legality issues over two decades ago to try and counter the negative impacts of illegal practices in tropical forests on the attainment of the Organization's objective to promote SFM. ITTO's approach to improving forest governance in the tropics has four interlinked components as shown in Figure 1.

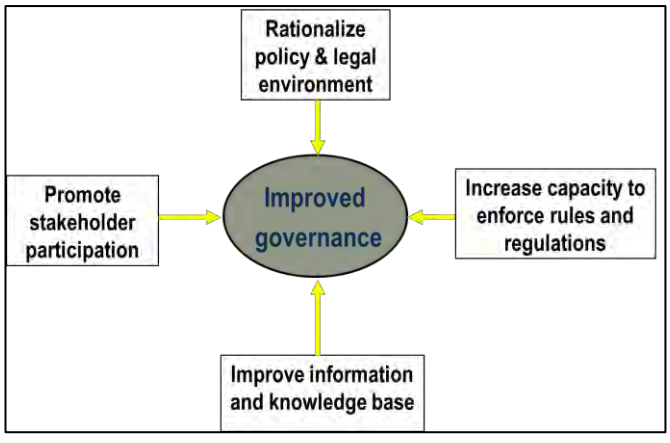


Figure 1. *Components of improving forest governance.*

The following sections provide an overview of recent ITTO activities under each of these components.

1.1 Rationalizing policy/legal environment

ITTO's has undertaken case studies on forest law enforcement (FLE) and illegal trade in many countries. These studies have found that conflicting laws and/or incoherent policies can be a significant factor in contributing to illegality in the forest sector. ITTO projects in several countries have therefore assisted in identifying underlying causes of illegality and in drafting coherent, consistent, enforceable forest legislation. Country diagnostic missions have also helped to identify underlying problems and have promoted improved policies for FLE. More specifically, a dedicated program to improve implementation of CITES requirements for listed tropical tree species helps to ensure that forest laws are consistent with CITES regulations/requirements.

1.2 Building capacity

ITTO supports many capacity building measures to promote SFM in the tropics. Of specific relevance to its work on improving forest governance are the following initiatives:

- FLE Best Practices workshops with FAO
- Large training programs to:
  - Improve forest statistics
  - Promote use of tracking technology
- Promoting phased approaches to certification
- Promoting NGO/civil society involvement in forest monitoring
- Encouraging countries to engage with international initiatives (eg FLEGT) and in bilateral discussions/agreements, share experiences

1.3 Improving data and knowledge

ITTO's work to improve transparency and knowledge in the tropical forest sector contributes to better forest governance. This work includes:

- a bi-weekly market information service newsletter that provides price and trade information for a range of tropical timber species/products, helping to provide transparency and prevent tax/royalty evasion
- a biannual review of production and trade statistics that provides detailed information on trade flows
- trade discrepancy studies and production/capacity/materials balance comparisons
- a project to independently monitor FLEGT-licensed timber entering the EU market
- timber tracking projects in many countries
- publication of the "Tracking Sustainability" report, providing descriptions of tracking technologies, case studies on their use, and recommendations for tropical countries on their application
- several projects using satellite imagery together with geographic information systems (containing details of approved concessions, roads, etc) to spot illegal forest clearing and track legal timber (see Figure 2).

Figure 2. *Detection of forest clearing outside an approved concession using IKONOS (4 m) and Landsat 5 (30 m) satellite data in Guyana.*

1.4 Promoting stakeholder involvement

The involvement of local people and communities living close to (or in) the forest is often essential to exposing and solving problems of forest governance. ITTO promotes stakeholder involvement by working with its Trade Advisory Group (TAG) and Civil Society Advisory Group (CSAG) in a number of ways, including:

- support for civil society – private sector partnership grants to contribute to SFM and verifiable legality/certification in many countries
- convening international conferences arising from recommendations of TAG/CSAG Panel on Illegal Logging /Illegal Timber Trade on topics including:
    - timber transport
    - indigenous/community forestry
    - tropical forest tenure

2. ITTO support mechanisms

ITTO support is channeled to countries through a number of different funding windows. These include a regular project cycle that allows countries to submit projects twice a year, thematic programs dealing with topics approved by ITTO's governing Council and other programs and activities under the Organization's biennial work programs. Two of the most important funding mechanisms relevant to ITTO's work on forest governance are the thematic program on tropical forest law enforcement, governance and trade (TFLET) and a program to assist countries to implement CITES listings of tropical tree species.

2.1 TFLET

- Tropical Forest Law Enforcement, Governance and Trade Thematic Program since 2008
- Rationalizes ITTO's work, provides dedicated funding window, one of four thematic programs approved/funded under International Tropical Timber Agreement 2006

- Over $10 million distributed to 50 projects in 25 countries to date; main themes timber tracking and community empowerment

2.2 CITES Program

- Assists countries to implement CITES provisions for listed tropical tree species
- Over $15 million from multiple donors (two-thirds EU) since 2007, 70+ projects in main exporting range states focusing on:
    - targeted inventories/management plans
    - non-detriment findings (NDFs) of sustainable production required for export of CITES Appendix II listed species
    - training/capacity building on CITES implementation
    - tracking of products covered by NDFs

Table 1 lists recent ITTO support under these funding windows for relevant projects and activities. Details on all of these projects/activities (including all reports and other outputs) are available on ITTO's website (www.itto.int).

3. Conclusions and lessons

Through its work on forest governance (specifically timber tracking), ITTO has reached the following conclusions and learned the following lessons:

- Timber tracking systems (TTSs, which exist in some form in most countries) are increasingly relevant for demonstrating legality and meeting market requirements (e.g. FLEGT VPA, U.S. Lacey Act, etc).
- For most tropical countries already involved in forest certification or monitoring species covered by international regulations such as CITES, chain of custody monitoring including TTSs already in place or planned. Such systems are deemed essential for the achievement of sustainable forest management (SFM) although technology cannot replace the human capacity necessary for SFM.
- Technology levels used must be appropriate to each individual country/industry and adequate capacity building needs to be undertaken to ensure sustainability and local ownership of the system after any pilot phase.
- Technologies such as DNA and stable isotope analysis can help to verify the accuracy of information generated by TTSs and support SFM. Support for establishment of TTSs and capacity building (esp. for smallholders) remains a necessity.
- Population assignment (back to forest or concession area) preferable to individual tree assignment especially where chain of custody is weak/incomplete.
- Acquisition of samples continues to be a challenge in many tropical countries, particularly for CITES listed species where there is a need for clarity in regulations regarding export of research specimens.
- Discussions required on benefit sharing to ensure DNA and other molecular data can be equitably shared.
- Continued work is required to consolidate reference sample databases to ensure wide availability and application to realize the potential of DNA based timber tracking and identification technologies. The Global Timber Tracking Network (GTTN), of which ITTO is a founding member, is a promising initiative in this regard.

Table 1. *Recent ITTO-funded tracking/identification work.*

Title	**Executing Agency**	**Species**
Developing DNA database for *Gonystylus bancanus* in Sarawak	Forestry Department Sarawak/Sarawak Forestry Corporation	*Gonystylus bancanus*
The development of *Gonystylus* spp. (ramin) timber monitoring system using radio frequency identification (RFID) in Peninsular Malaysia	Forestry Department Peninsular Malaysia	*Gonystylus* spp.
Use of DNA for identification of *Gonystylus* species and timber geographical origin in Sarawak	Sarawak Forestry Corporation	*Gonystylus* spp.
Training interested parties on the verification of CITES permits and the use of the "CITES Wood ID" in DRC	MECNT (Division for Wildlife Resources and Hunting)	*Pericopsis elata*
Pilot implementation of a DNA traceability system for *Pericopsis elata* in forest concessions and sawmills in Cameroon and Congo	Double Helix/ ANAFOR/CNIAF	*Pericopsis elata*
Pilot Implementation of a DNA traceability system for *Prunus africana* in *Prunus* Allocation Units in Cameroon and Democratic Republic of Congo (DRC)	Double Helix/ MINFOF/MECNT	*Prunus africana*
Using the Near Infrared Spectroscopy (NIRS) technique on a pilot scale, as a potential tool for the monitoring of mahogany trade	Brazilian Forest Service Forest Products Laboratory (SFB/LPF)	*Swietenia macrophylla* (mahogany) *Carapa guianensis* (crabwood/ andiroba) *Cedrela odorata* (cedar) *Micropholis melinoniana* (curupixá)
Establishment of a forensic laboratory for timber identification and description in the implementation of legal proceedings and traceability systems for CITES listed products (Guatemala)	Fundación Naturaleza para la Vida –FNPV (Nature for Life Foundation)	*Swietenia macrophylla* *S. humilis* *Dalbergia calycina* *D. retusa* *D. tucurensis* *D. stevensonii* *Guaiacum* spp.
Establishment of a fully documented reference sample collection and identification system for all CITES-listed *Dalbergia* species and a feasibility study for *Diospyros* and look-alike species	Institute of Integrative Biology (IBZ), Switzerland	*Dalbergia* *Diospyros* spp.(Madagascar)
Development and implementation of a species identification and timber tracking system in Africa with DNA fingerprints and stable isotopes (PD 620/11 Rev.1 (M))	Thünen Institute of Forest Genetics	Iroko *(Milicia excelsa, M. regia)* sapelli *(Entandrophragma cylindricum)* ayou *(Triplochiton scleroxylon)*
Implementing a DNA Timber Tracking System in Indonesia (TFL-PD 037/13 Rev.2(M))	University of Adelaide, Australia	Red meranti group Light red meranti

Chapter 1.4.

The Global Timber Tracking Network

Jo Van Brusselen
European Forest Institute, Finland

Abstract

The Global Timber Tracking Network (GTTN) promotes innovative tools to verify trade claims of wood-based products. GTTN intends to assists global action against illegal logging and related trade, in support of the work of e.g. Monitoring Organisations, law enforcement agencies, or to assist timber traders and operators' own due diligence systems. GTTN therefore brings stakeholders together to improve development and sharing of reference data; standardization of methods; development of a laboratory directory; networking, communications and advocacy activities.

Keywords: FLEGT, GTTN, timber tracking, wood products trade, tree species verification, verification of geographical origin

1 Introduction

The objective of the Global Timber Tracking Network (GTTN) is to promote the operationalization of innovative tools to verify trade claims of wood-based products. In doing so, GTTN intends to assists global action against illegal logging and related trade, in support of the work of e.g. Monitoring Organisations, law enforcement agencies, or to assist timber traders and operators' own due diligence systems. In a wider sense timber tracking can support global forest governance.

A scientific proof of tree species and geographic origin of a wood product, can help determining whether wood originated from places where logging was allowed, or not, e.g. from outside a forest concession zone. This is possible by holistic application of methods involving e.g. wood anatomy (including neural network-based multispectral macro- and microscopic imaging), genetics, stable isotope analysis, mass spectrometry, and near-infrared (NIR). This can support trading timber with the confidence that tree species and origin match expectations.

GTTN is intended as a global platform bringing together scientists, policy makers and other key players. GTTN is an open alliance that cooperates along a joint vision and the network activities are financed through an open multi-donor approach. The GTTN secretariat activities are financially supported by the German Federal Ministry of Food and Agriculture (BMEL).

***Figure 1**¡Error! Marcador no definido.**. Participate in a growing global network. Status of 20170901. The network remains open for registrations: for developers, providers and users of timber tracking, but also for other interested stakeholders.*

2 Cooperation through working groups

The key objectives for the Global Timber Tracking Network in phase 2 remain to further develop and expand the network, to seek new partnerships both with (potential) providers as well as with (potential) users of timber tracking services, to facilitate active collaboration, coordination of activities and to advocate for funding with the donor community.

While GTTN is maintained by a small secretariat to coordinate activities, the knowledge and know-how that will make GTTN strong, relies in a network of active members. At the beginning of phase 2, the network opened up for additional members with an on-line survey. This resulted in the identification of over a hundred specialists, who agreed that cooperation is important to reach shared aims, and who indicated commitment to contribute to the activities of three working groups, one for each activity strand. The survey remains open and accessible via the website. www.globaltimbertrackingnetwork.org.

3 GTTN key objectives and activities in phase 2

GTTN seeks collaboration through three activity strands: (i) seek agreement on international standards for sampling and reference data development and use, (ii) service portfolio and lab-finder tool, and [meta]database of reference data, and (iii) communication and advocacy.

3.1 International standards

The GTTN project seeks to support the development of timber tracking methods that would enable reliable identification of timber type and origin and would be used for combating illegal timber harvesting and trading. However, to achieve this overall objective international standards and methods should be identified or developed and agreed upon. These standards should include procedures for sampling, material storage and documentation, material exchange, test sample preparation, method application, data analysis, lab accreditation and applicable regulation.

In addition to the work on standards and methods, it will be important to improve understanding on the potential role of tracking technologies in the practical implementation of legality verification policies through analysing implementation practices, and exploring the demands of different concerned stakeholders (authorities, industry, civil society, science) and connecting stakeholders' demands and technological possibilities through analysis and networking in view of possible standardization processes.

3.2 Service portfolio and lab-finder

A service portfolio and lab-finder will be developed to provide interested external users with information on which methodologies are available to check a specific trade claim and whom can be contacted to perform the testing.

These external users will most likely be either interested from the perspective of due diligence such as forest-based industries, monitoring organisations, or then from the perspective of law enforcement.

The availability of expertise will be influenced by the type of product (e.g. full wood, fibre board or paper), by trade claim i.e. declared tree species and/or geographic origin, and also by the location and type of the service user, and what the result would be used for (e.g. a service provider of a result to be used in a court of law will need to fulfil certain requirements).

3.3 [Meta]database of reference data

An on-line reference data needs to be further developed and operationalized to provide internal users with a secure access to a data repository with reference data to identify species and or geographical origin of wood fiber. The key focus will be reference data for (geo-referenced) DNA, stable isotope data, while inclusion of wood anatomy catalogues, mass spectrometric data ought to be considered, amongst other.

Optimally users would have put their data into the database, signed a data sharing agreement, successfully participated in ring tests for standardisation and are ready to provide the lab services. More likely the system will need to cater differentiated access levels for different types of registered users.

The reference data system will need to cater for varying requirements from reference data producers that range from researchers working on a doctoral degree without having back-up systems on the one hand, to expertise centers and companies that have in-house systems

already developed. Some data holders will not want their reference data to be stored elsewhere for the risk of losing metadata, i.e. descriptive information concerning the reference data and possibly the samples that the reference data were created upon. The capability for durable storage of reference data together with the appropriate metadata may be very limited for individual researchers however, who would be on the requesting side for a central data repository.

The implication would be that the reference data system would need to be centered about the development of a set of metadata that would answer to the requirements of reference data holders, while also giving an option for storage of metadata for those who do not have the capacities themselves.

3.4 Communication and advocacy

GTTN intends to enable its stakeholders to come together to cooperate and exchange ideas and information and to inform researchers and stakeholders about available services. GTTN stakeholders get together in two global working group meetings and three regional events (Africa, Latin America, and South-East Asia). GTTN phase 2 organises one working group meeting in 2017 and one in 2018; two regional meetings in 2018 and one in 2019.

Whether it concerns the GTTN website, news, databases, guidelines development, workshops, the success of GTTN communication and advocacy will benefit from the cooperation and inputs of the GTTN members. GTTN will also take a pro-active stand to help members with outreach activities, news, success stories, and event announcements.

An important tool in support GTTN advocacy will be the development of a strategic research agenda, which will identify urgent needs for timber tracking research, sampling, methods development, infrastructure and capacity building.

The advocacy function of GTTN will also consider the important issue of intellectual property rights (IPR), access and benefits sharing (ABS), as relating to the Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the convention on biological diversity.

4 Seeking value in partnership

The success of the Global Timber Tracking Network will finally depend on the endorsement and uptake of timber tracking services by users. However that goal is preceded by the need for active cooperation by science and service communities, on standard setting, data sharing etcetera. The best way to achieve successful cooperation is by making sure that effort put into cooperation will result in something useful for all parties engaged.

Results or benefits from cooperation for researchers and timber tracking service suppliers might entail e.g. an increased visibility of their work via events, website, publications; the broaden of their network, giving a possibility to synergize efforts; ability to reach broad consensus on standardization; access to reference data could help diversify service portfolios; international engagement can help to raise a laboratory’s profile; new ideas might result in new projects.

The Global Timber Tracking Network wants to encourage (potential) users of services to be involved in the project, e.g. to influence the way in which they want to find information through GTTN; to help prioritizing the tree species that collection of new reference data is most urgent for; or even to get actively involved in the collection of sample material, which in turn would lead to a higher precision of timber tracking services that they would be in need of. Users' engagement can be made visible through GTTN communication tools, which in turn helps users promote their commitment to verified legal supply chains.

Last but not least, GTTN should remain open to contribute to and benefit from other meetings, such as the OECD-CRP sponsored workshop. While the workshop focused on rapid through-put genetics analysis, the recommendations from the workshop report will be well-worth the consideration of the wider GTTN network.

Conclusions of PART 1

Economical, legal and environmental aspects of the timber illegal logging and trading

A prioritized list of forest tree species should be prepared based on a reviewed analysis of GTTN previous list, including updated high value traded, very endangered CITES listed species and species described in the frame of bilateral agreements (i.e. FLEGT).

- - - - - - -

Forest tree traceability should be a multidisciplinary effort, including active participation of taxonomists for the identification of reference samples, and sampling design scheme when these samples are not included in existing collections. Arboretums and botanical gardens will be key actors in the selection of the reference samples.

- - - - - - -

The support of origin countries, as well as international organizations (i.e. CGIAR Centers, INTERPOL, etc), is crucial to warrant efficient sample collection efforts. MTAs should be standardized to ensure recognition of providers and traditional knowledge. Access and benefit sharing should also be developed to guarantee benefits stay in the origin countries.

PART 2

HIGH-THROUGHPUT TECHNOLOGIES TO IDENTIFY TREE VARIABILITY

Chapter 2.1.

A set of genomic and electronic resources to discriminate genetic units in European white oaks

Antoine Kremer
INRA, UMR 1202, France

The evolutionary history of white oaks in Europe has been and still is a shared matter of research within the community of forest geneticists across the continent. During the past three decades, collaborative research projects have been implemented along this research area, and benefited of financial support by the European Union, and by national funding agencies. The rationales of these projects were rooted in basic and applied research goals as: (1) reconstruct evolutionary trajectories of oak species in Europe since the last glaciation, (2) identify key ecological, genetic and demographic mechanisms shaping extant genetic diversity in oak forests (3) provide a fine scale description of the geographic and spatial distribution of genetic diversity based on genomic data.

Among other outcomes, the development of molecular tools allowing to differentiate genetic units which are of interest in forestry forensics (oak species, populations) was a major deliverable of these collaborative projects. Concerning the species level, it is well known that European white oak species can hardly be identified with wood anatomical traits, while other morphological traits may not always be accessible on log piles or trucks. In addition, white oaks are widely interfertile, adding to the complexity to identify discriminant genetic or morphological signatures. Concerning the population level, the search for traits or markers varying between populations was constrained by the large scale and long distance pollen flow with tends to homogenize populations throughout the landscape. Luckily, oak seed are less dispersed and they are the only vector of chloroplast transmission; hence the spatial distribution of chloroplast genetic diversity is mainly shaped by natural or human mediated seed or seedling (plantation) transfers. This reasoning led us to explore extensively genomic variation within the chloroplast genome, and resulted in a comprehensive geographic map of chloroplast DNA genetic types (haplotypes) across Europe. These data have been assembled in a public web accessible data base (http://gd2.pierroton.inra.fr/), that is steadily populated with new published results. To sum up, three decades of research in population and evolutionary biology in the two European temperate white oak species (*Q. petreae* and *Q.*

robur) resulted in two tools that can be applied to forest forensics in the context of combating illegal logging:

- A set of genomic markers allowing to differentiate the two species
- A georeferenced data base of chloroplast haplotypes, with a fine scale coverage across the continent

Genomic tools for species and population differentiation

Ever since molecular markers have been used in population genetics, attempts to discover species specific markers in European oaks have been implemented. These attempts resulted in congruent conclusions showing that such markers are extremely rare and widely distributed throughout the genome. It is out of the scope to recall the multiple genetic surveys that were conducted with different marker sets during the past two decades. The recent accessibility to a reference genome has confirmed these expectations by mapping full genome sequences of different white oaks on the reference genome. For the time being, in the case of *Quercus petraea* (sessile oak) and *Q. robur* (pedunculate oak), which are the two most economically important and widespread species, a set of 16 SNPs allows to obtain distinct separation between the two species. These have been tested on independent maker sets.

Concerning the population level, research efforts have concentrated on the organelle genomes, as mitochondria and chloroplast are transmitted only by seed and exhibit strong population differentiation. Indeed, most populations (forests) are entirely fixed for a given haplotype. About 42 haplotypes were identified across Europe; with 8 showing very large geographic distribution, within three longitudinal zones (Atlantic Zone, Central and Eastern Zones). Originally the genomic assay to identify the haplotypes was a PCR RFLP technique. But recently this assay was upgraded into a SNP detection of 35 SNPs. Extraction techniques to recover chloroplast DNA from wood were also tested and proved to be successful especially when extractions are conducted on sapwood.

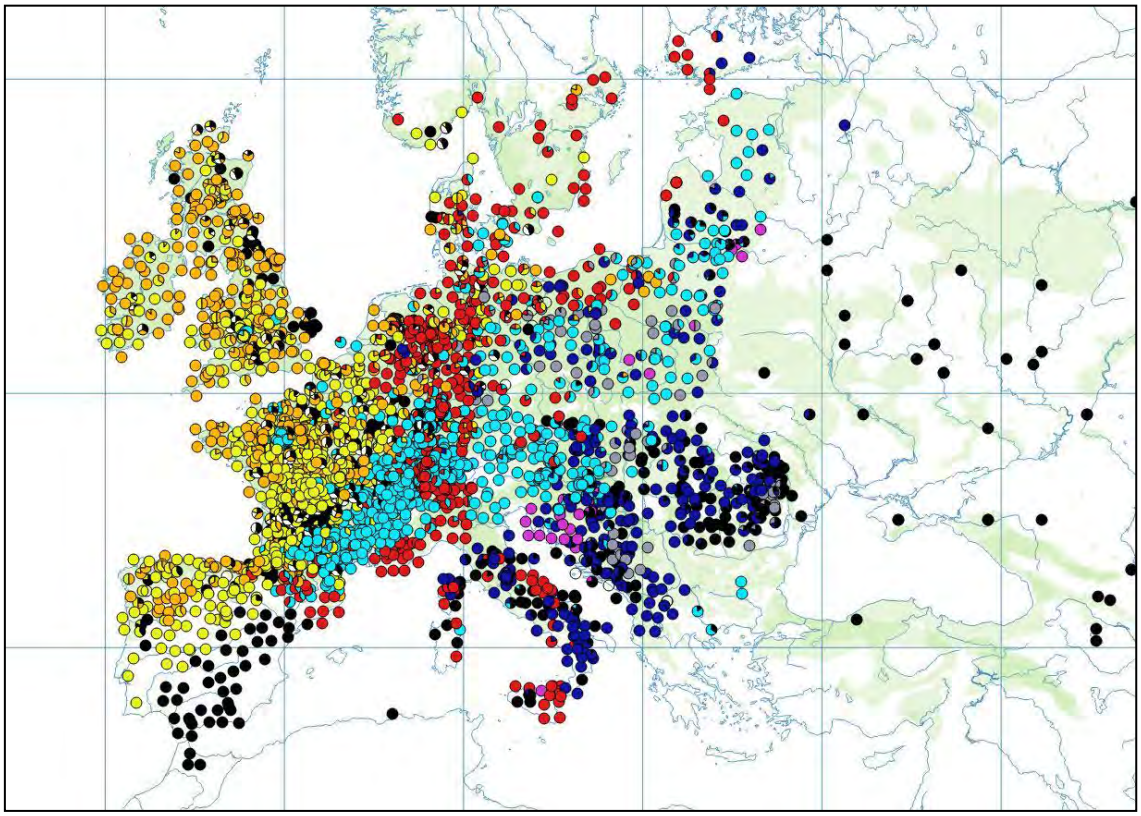
The following table summarizes these achievements and indicates the core publications where the molecular and genetic techniques are described.

Table 1. *Molecular assays used routinely to differentiate species (Q. petraea /Q. robur) and populations. References are added where the methods are described*

	Species differentiation Nuclear genome	**Population differentiation Chloroplast genome**
"Traditional "genotyping	10 to 15 microsatellites (2 multiplexes)	PCR-RFLP 4 cpDNA fragments
Reference	*Guichoux et al. 2011, Molecular Ecology Resources 11: 578-595*	*Petit RJ et al. 2002, Forest Ecology and Management 156:5-26*
Standardized method	16 SNPs *Guichoux et al. 2013, Molecular Ecology 22: 450-462*	35 SNPs *Guichoux E, Petit RJ 2014, Patent DI-RV-13-00566*
References	*Truffaut et al, 2017 New phytologist 215:126-139*	

Electronic repository of georeferenced data of genetic diversity in oaks

As deliverable of the EVOLTREE network of excellence (http://www.evoltree.eu/), it was decided by the network to create a georeferenced data base of genetic diversity (GD^2) of European forest trees. GD^2 (http://gd2.pierroton.inra.fr/) contains passport data (geographic and ecological information) and genetic data (genetic arrays at the single tree level or at the aggregate population level for all different genetic markers) of tree populations, that were investigated in genetic surveys. The data base is regularly populated by published results as a continuous activity of the network. To this end, contacts are taken with authors of publications to share their data, if possible up to the single tree genotypes. The data base allows to visualize the distribution of genetic types (haplotypes, alleles, genotypes...) on different maps supports, and therefore allows to conduct any metanalysis of the distribution of genetic diversity across different geographical scales. It could therefore be used for tracing biological material (wood, or seed). To sum up GD^2 contains only published data that was afforded by authors willing to share their data. The figure attached visualizes an outcome of GD^2, concerning the chloroplast haplotypes of European white oaks (colors of the pie charts). For the sake of clarity this figure represents only the most frequent haplotypes. Within GD^2 the user has various options to sketch the distribution of haplotypes, by focusing either on specific areas, or limiting the analysis to certain haplotypes.



http://gd2.pierroton.inra.fr/

Finally the GD^2 data base is connected to the Evoltree eLab (http://www.evoltree.eu/index.php/e-recources/elab) containing many other data bases with genetic and genomic information of trees (genetic maps, gene sequences, provenances, etc...) through a standardized HTTP transmittable interface, so that queries can be made

Chapter 2.2.

Towards the development of a traceability system for Canadian forest products

Nathalie Isabel, Julie Godbout, Claude Bomal, Miranda Williamson, *and* ***Ken Farr***
Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Canada

Background

It is a fact: consumers, retailers, investors, and governments want quality products that are both safe and delivered through legal and sustainable practices. The solution put into place to meet those challenges requires the implementation of a traceability system, but what is traceability? Within a chain of custody, "traceability is the ability to trace the history, application or location of an entity by means of recorded identifications" (ISO 1994). Consequently, it aims at tracing back a product at every stage of the supply chain from the source to the market. In the context of forestry, traceability would essentially refer to the capacity to inform about the identity of a given entity (seed, plant, wood product) at the intra- or interspecific level or at the provenance level (geographic origin) (Godbout *et al.* submitted).

In a near future, traceability will likely play a more important role in removing technical barriers to trade or resolving some of the issues encountered by the Canadian forest sector. At the moment, this sector is in transformation and is facing many challenges, including issues related to legality, forest certification, and product safety. The objective of this paper is to present a short overview of the opportunities offered by the genomics tools and resources that have been developed for many Canadian commercial forest species to provide traceability solutions. We briefly define the concept of traceability and describe how genomic tools can be brought into play within a traceability system with concrete examples specific to the Canadian forest sector. Thus, we present how genomic tools allow us to answer questions related to traceability. Finally, we conclude by presenting the challenges related to the development of such systems. We currently have a manuscript under revision that provides a lot more details and examples on how genomic tools could contribute to the development of traceability systems in the forest sector.

Canada's boreal forest ecosystem context at a glance

Canada's forests cover 347 million hectares, of which provincial governments and territories own 77% (The State of Canada's Forest 2016). The boreal forest ecosystem

dominates the landscape and spruces are the most abundant species. The maintenance of ecosystem health and diversity is largely influenced by natural disturbances, such as fire and insect outbreaks, that also vary in terms of recurrence and severity. In 2016, this represented 7% of the country's forest (The State of Canada's Forest 2016). Besides, less than one million (>750 000) hectares of forest were harvested, which represents 0.3% of the 347 million hectares. Fire, insects, and diseases are the main drivers of natural regeneration of the forest. Most of the forested areas harvested are located on public land and they must be regenerated either naturally (40%) or artificially (60%) to continue providing ecosystem services and producing wood fibre. Forest regeneration after harvesting activities are monitored and regulated by the provinces. Forest products contribute to nearly 7% of all Canadian exports and the demand for certified forest products has grown rapidly as consumers seek out sustainable products. Certified forest products and forest certification allow the forest products industry to communicate to consumers their commitment to sustainable forest management practices. The criteria needed to achieve forest certification are diverse and varied, but maintaining certification is essential to ensure access to large overseas markets. Canada actually has 166 million hectares of forest certified by a third party (The State of Canada's Forest 2016).

Traceability in an ideal world

Until recently, the majority of traceability systems used, as for certified forest products, have relied on paper permits and stamps, which might be counterfeited. An ideal traceability system would rely on robust and reproducible tools (e.g. Lowe *et al.* 2010, 2015) that would essentially enable tracing back a wood product, at various levels of resolution corresponding to different genetic units: from a unique fingerprint specific (e.g. clonal variety) to a group of species of the same genus (Godbout *et al.* submitted). Although more challenging to obtain, it would also make possible to designate the geographic origin (provenance) of wood products.

Over the last decade, major investments have been made in genomics, which has led to the development of abundant genomic resources for Canadian commercial forest tree species (e.g. Pavy *et al.* 2013). This has created opportunities for developing or at least testing the possibility of developing a traceability system along the chain of custody. Although genomics offers us powerful tools to answer traceability issues, it is critical to specify the ultimate goal of a project in order to define the question(s) to be answered (Godbout *et al.* submitted). This predefined question will largely influence the degree of precision required (one genotype, multiple species, geographic origin) which in turn provides guidance about the choice of technoscientific tools to be used and the depth (richness) of the reference database to be developed. Since the technoscientific tools and the reference database are highly dependent on each other, equal importance should be given to their development.

Within the forestry context, three different levels of resolution, which delineate genetic units, can be envisioned for the development of a traceability system (Godbout *et al.* submitted). A first level would inform about the identity of a product, such as the unique fingerprint generated to distinguish different genotypes within a species. A second level of resolution would consist in developing sets of species-specific markers that allow distinction between species. Finally, the last level and not the least, would provide enough precision to determine the geographic origin of a product.

Unique fingerprint (intraspecific level)

We illustrate the first level of resolution by presenting a simple methodology that made it possible to develop a traceability system by relying on SNP-based markers, which is used, in this particular case, as a quality-control assurance system for the large-scale production of high value seedlings (Godbout *et al.* 2017). Genomic resources were already available in white spruce (e.g. Pavy *et al.* 2013). In Eastern Canada (Québec province), one of the strategies for large-scale production of white spruce seedlings of elite varieties is through the somatic embryogenesis (SE) process. Tissue culture used in combination with vegetative propagation allows to capture 100% genetic gains (additive and non-additive), but may sometimes imply that an error occurring somewhere in the process, could also be multiplied.

For the development of this traceability system, two objectives were set: 1) to develop methods that make it possible to trace back the origin of the seedling produced (cross of origin), and 2) to generate a unique genetic fingerprint that could be used to differentiate each embryogenic cell line, from the cryobank to the field. Two additional criteria were also identified: a minimum number of low-cost DNA markers should be used (40 SNPs), and the traceability system (from DNA extraction to data interpretation) should be easy to use by non-specialists. The 40 SNPs corresponded to the capacity of a single Sequenom® genotyping array. Accordingly, this traceability system was developed using simulated data sets and classification methods currently available in the literature (Godbout *et al.* 2017). In summary, 73 parents, that had already been genotyped for nearly 500 SNP markers, were used to generate these data sets, that simulated crosses representative of the white spruce breeding program. From the simulated data sets, we selected the 40 most informative markers using three discrimination procedure (F_{ST}, MAF, and Random Forest; Breiman 1999) and two assignment methods were compared (FAP, Taggart 2007 and PAPA; Duchesne *et al.* 2002). Then an array of 40 selected SNPs was designed and genotypes were obtained for 4346 samples (2845 trees and 1501 tissues) representing 1517 cell lines. The rate of misidentification was estimated and various sources of mishandling or contamination were identified (for details see Godbout *et al.* 2017). Most importantly, a unique fingerprint was generated for a majority of the embryogenic cell lines. This quality-assurance traceability system is currently used by the nursery. Therefore, we consider that the proposed approach and methods can be easily utilized to develop quality-assurance control systems for any other living production systems for which genomic resources are already available.

Tracking exotic alleles in natural populations (interspecific distinction)

Intensively managed plantations of fast-growing trees, such as poplars, are seen as a way to meet fibre production quotas on smaller plots of land. Novel tree varieties with superior growth or disease resistance are often selected to ensure maximum yield. These novel varieties, hereafter named exotic hybrids, are derived from breeding native trees with exotic species. The *Forest Stewardship Council (FSC*) stipulates that “The use of exotic species shall be carefully controlled and actively monitored to avoid adverse ecological impacts”. Exotic hybrids in plantations can contaminate nearby populations of native trees through pollen or seed dispersal. This contamination can result in the flow of exotic genes into native

populations through the process of introgression. Escaped exotic genes would violate FSC criteria and potentially threaten the forest certification of Canadian forest products.

In this case, we have developed a traceability system to distinguish poplar species among them (second level of resolution) in order to examine the risks exotic genes pose to Canada's native forests. Using modelling approaches and technoscientific tools (Meirmans *et al.* 2007, Talbot *et al.* 2011, Isabel *et al.* 2013), that allow to track gene flow from exotic plantations (including windbreaks and urban trees) to natural populations, we found that: 1) some native tree species are more likely to experience exotic gene contamination (Meirmans *et al.* 2010); 2) small populations of native species located near plantations are more at risk of contamination (Meirmans *et al.* 2010, Thompson *et al.* 2010); 3) disturbed habitats provide a niche for exotic hybrids to establish themselves (Thompson *et al.* 2010); 4) the fate of exotic genes in native gene pools is determined by the advantage it provides (Meirmans *et al.* 2009); 5) newly formed hybrids are not necessarily superior to their parental species (Roe *et al.* 2014a, b); and 6**)** the gender of novel varieties of poplar used for plantations must be taken into consideration in management strategies to minimize the occurrence of crossing. Male poplar varieties should be preferred to female trees since the exotic pollen coming from plantation is diluted in the native pollen cloud (Talbot *et al.* 2012, LeBoldus *et al.* 2013). We also showed that effective monitoring and risk assessment is imperative to ensure plantations with exotics do not threaten the genetic integrity of native forest trees or Canada's forest certification. Tools and protocols for monitoring were adopted and put into place with the end-users (tree breeders and forest companies).

Geographic origin (third level)

We are currently testing the feasibility of using genomic tools to address the question of traceability for two commercial forest tree species present in Canada. We want to take advantage of existing genomic resources and genotype data sets from many populations across the species' range (thousands of SNPs and hundreds of individuals) and to test the ability to use SNPs to identify the geographic origin of "unknown" samples. This proof-of-concept project started recently.

Challenges

In order to establish trust among the various stakeholders and consumers throughout the chain of custody, a traceability system needs to demonstrate its **independence** (audits by a third party) and **reproducibility** (standardized protocols and validated data using reference samples) (Godbout *et al.* submitted). To date a majority of traceability systems, like those used for certified forest products, relied on paper permits. To facilitate the creation of valuable and transparent traceability systems based on DNA tools, that can be trusted because they are less prone to falsification, there is an urgent need to address the following issues: 1) open data and ownership, because most of the data already available for a large number of forest species was generated using public fundings; 2) both the location and the access to collection of reference samples with standard operating protocols and; 3) proof-of-concept studies for species of interest.

A final and essential element to consider for the development and implementation of a traceability system in forestry is the relationship to be established between the end-users and scientists (Godbout *et al.* submitted). This link is a prerequisite to the creation of a traceability system that will meet the needs of the former while using the technical and scientific skills of the latter. To this end, the development of a project, i.e., determining the question to be answered, the methods tested and the material to be used, should ideally be done through discussions between both parties. Such approach will foster the use of traceability system by users.

References

Breiman, L. 1999. Random Forests. Machine Learning 45, 1-35.

Duchesne, P., Godbout, M.-H., and Bernatchez, L. 2002. Mol. Ecol. Notes 2, 191-193.

Godbout, J., L. Tremblay, C. Levasseur, P. Lavigne, A. Rainville, J. Mackay, J. Bousquet, and N. Isabel. 2017. Front. Plant Sci. 8: 1264.

Isabel, N., M. Lamothe, and S.L. Thompson. 2013. Tree Genet. Gen. 9: 621-626.

[ISO] International Organization for Standardization. 1994. ISO/TC 176/SC 1, Geneva, Switzerland.

LeBoldus, J, N Isabel, K D Floate, P Blenis, and BR Thomas 2013 Plos One 8:e84437

Lowe, A.J., K.-N. Wong, Y.-S. Tiong, S. Iyerh, and F.-T. Chew. 2010. Silvae Genet. 59: 263-268.

Lowe, A., E. Dormontt, and A. Rimbawanto. 2015. Austral. Cent. Intl. Agric. Res. Final Rep. FR2015-15.

Meirmans, P.G., M. Lamothe, P. Périnet, and N. Isabel. 2007. Can. J. Bot. 85: 1082-1091

Meirmans P.G., J. Bousquet, & N. Isabel. 2009. Evolutionary applications. DOI 10.1111/j.1752-4571.2008.00050.x

Meirmans, P.G., M. Lamothe, M.-C. Gros-Louis, D. Khasa, P. Périnet, J. Bousquet, and N. Isabel. 2010. Am. J. Bot. 97: 1688-1697;

Pavy N, F Gagnon, P Rigault, S Blais, A Deschênes, B Boyle, B Pelgas, M Deslauriers, S Clément, P Lavigne, M Lamothe, J Cooke, JP Jaramillo-Correa, J Beaulieu, N Isabel, J Mackay, J Bousquet. 2013. Mol Ecol Res 13: 324;

Roe A, C MacQuarrie, M-C Gros-Louis, J Lamarche, T Beardmore, S Thompson, P Tanguay, N Isabel 2014a,b. Ecol. Evol. 4: 1876;

Taggart, J.B. 2007. Mol. Ecol. Notes 7, 412-415; Talbot, P., S.L. Thompson, W. Schroeder, and N. Isabel. 2011 Can. J. For. Res. 41: 1102-1111.

Talbot, P.; Schroeder, W.R.; Bousquet, J.; Isabel, N. 2012. Talbot, P.; Schroeder, W.R.; Bousquet, J.; Isabel, N. 2012. For. Ecol. Manag. 285:142-152;

The State of Canada's Forests*: Annual Report* 2016. *Natural Resources Canada;*

Thompson, S.L., M. Lamothe, P.G. Meirmans, P. Périnet, and N. Isabel. 2010. Mol. Ecol. 19: 132-145.

Chapter 2.3.

Use of genotyping technologies to identify tree variability

Valerie D. Hipkins
National Forest Genetics Laboratory, US Forest Service, USA

The US Forest Service, National Forest Genetics Lab (NFGEL)

The USDA Forest Service established the National Forest Genetics Laboratory (NFGEL) within the management branch of the Agency in 1988 to provide forest managers with the means to evaluate the genetic consequences of management actions. Lab work supports the conservation and management of all plant species, specifically to (1) identify sources of seed for species and population re-establishment, (2) assist with the breeding and production of new genotypes and seed sources for assisted migration in the face of changing climate, (3) identify endangered and invasive species, (4) identify species, populations, and communities that are sensitive to increased disturbance, and (5) track the movement of genetic resources.

NFGEL works closely with land managers, including law enforcement, to provide them key genetic information that is relevant and timely for their decisions. Society's ability to establish and sustain healthy forests and rangelands, especially in face of current pressures such as habitat fragmentation, climate change, and degraded ecosystems, requires an understanding of genetics. Information about genetics helps assess past, current and future biological changes, and provides implications for management options in the future. NFGEL uses state-of-the-art technology to address genetic conservation and management of all plant species using various laboratory techniques including DNA analyses.

Forest Service managers need scientific information about genetic structure to withstand scientific and legal challenges in ongoing and future negotiations, adjudications, and in Forest planning and project implementation. Application of science-based technology is essential for the Forest Service to meet its Organic Act purpose to effectively manage National Forest System lands to secure multiple resource benefits. Organizational effectiveness requires capturing the benefits of rapidly evolving technology. Effective use of new science and technology requires guidelines that have passed scientific peer review. An essential part of our effort is to develop reliable, effective, low-cost, time-efficient technologies for characterizing genetic variation in all plant species to aid in adaptive planning efforts on forest and

rangelands throughout the Nation. Developing standard approaches for inventorying and describing shifts in genetic variability over spatial or temporal scales is also vital. Development includes using new computer and laboratory technologies to extend our capability to perform analyses, provide diagnostic and design assistance, and transfer technology to field specialists.

NFGEL uses a variety of genetic markers to assess variability. Markers are chosen foremost for their effectiveness at addressing project and management objectives. Secondary considerations are marker availability without the need for additional development, low cost, and fast application. When appropriate, we will assess variation with some of the newest technologies, and at other times we use some of the oldest markers, such as isozymes. The most commonly used markers for assessing diversity in the lab are still simple sequence repeats (SSRs) or microsatellites. Although SSRs are still extremely useful markers, the trend is to develop and use SNPs as discovered through high-throughput sequencing technologies. The use of high-throughput sequencing techniques, also called next-generation sequencing (NGS) technologies, to identity genetic variability can provide powerful solutions for species and source level identification. These technologies may provide practical solutions to the challenge of dealing with tree species that are often characterized by many large populations that possess complex and large genomes more than 10 Gb in size (Eckert *et al.* 2009). On some NFGEL projects, we are using unique SNP markers to identify differences among genotypes for studies of genetic diversity and population structure studies of forest trees.

Variability Assessments

NFGEL's interest in high-throughput technologies arises from needing to assess relevant levels of variation useful for characterizing species (taxonomy), population (source), and sometimes individuals, often in tree species that are not well characterized. Although species conservation and restoration drive much of our work, the same study designs are used for our timber tracking efforts. Workflow for many of these projects includes: (1) characterize the level and structure of genetic variation in the species &/or range of interest (this serves as the reference database or is used as the reference material), (2) chose appropriate markers to build the database that are robust and reliable for genotyping, contain sufficient levels of variation so that they are able to discriminate samples and be stable enough to achieve compatible data among samples over time and between labs, and (3) obtain usable DNA of sufficient quality and quantity from difficult samples (either from vegetative tissue from trees located in remote locations, or from wood).

Genetic variability in Golden Chinquapin (Chrysolepis chrysophylla)

Golden Chinquapin is an evergreen tree or shrub endemic to the western United States and is related to the beech family (Fagaceae). When growing to its full height as a tree it can reach 20-40 m tall. When classified as a shrub it grows between 3-10 m tall. Due to its 'sensitive' legal status in Washington, the level and distribution of genetic diversity must be characterized in the species for its successful management.

We created the range-wide reference database with two genetic markers: SSRs and SNPs. Sixteen SSR loci (Wu & Hogetsu (2009); Durand *et al.* (2010)) were run on 716 trees from 23 stands. Simultaneously, we also used 2b-RAD, a next generation sequencing approach for genome-wide genotyping (restriction site-associated DNA (RAD), based on sequencing

fragments produced by type IIB restriction endonucleases). Using this technology, we genotyped 628 trees at 2,021 SNPs with coverage ≥ 20X, with a mean 25% missing data per individual (range 10-62%). The SNP frequency over the entire data set is 2.05% (1 SNP every 48 bases). Compared to RAD and GBS (genotyping by sequencing), 2b-RAS uses a flexible reduction and has a faster bench protocol, though could have some challenges when applied to highly complex genomes.

We are currently evaluating the effectiveness of each dataset to characterize the genetic variation in the species, and are also looking at the markers practically (comparing the cost and time of development, the cost and time to run the markers, errors rates in the markers, and long-term stability of the datasets when adding new reference samples over time). In 2016, it cost approximately $20,000 USD to obtain the SSR database over a one-year period, and about $75,000 for the SNP database that took roughly 2.5 years to generate. Project co-operators include US Forest Service managers and University research geneticists.

Genetic variability in Mulanje Cedar (Widdringtonia whytei)

Mulanje Cedar is the national tree of Malawi. This species is declining due to over-exploitation, fire, and illegal harvesting. The species occurs naturally only in the Mulanje Mountain Reserve and is located in discrete regions, basins, and stands. Three offsite seed producing stands also exist, though they are of unknown origin. Objectives of the genetic work are to assess the diversity (level and structure) in remaining natural stands and in established seed orchards/plantations. Effectively, the goal is to build the reference database to use for restoration decisions, which could also be used for purposes of tracking illegal harvest/trade. Our partners include US Forest Service geneticists, Botanic Gardens Conservation International, Forestry Research Institute of Malawi, and Mulanje Mountain Conservation Trust.

Due to the need to obtain results within a one-year time frame, we are generating SSR data using approximately one dozen loci from *Widdringtonia cedarbergensis* found in “SSR markers from a thousand plant transcriptomes” (Hodel *et al.* 2016). There are a total of 850 samples, 700 of which are wood tissue from burned stumps. Isolating usable DNA from wood is a challenge for timber tracking efforts. Obtaining DNA of sufficient quality and quantity from wood samples will be necessary for the successful application of genetic marker technology and remains one of the primary challenges of these timber tracking efforts.

Concluding Remarks

Genetic tools are extremely useful for detecting species presence on the landscape, understanding the population structure and hereditary relationships of plants, and tracking plant material throughout a production supply chain. Efforts to achieve these goals would benefit from (1) a clearly defined management question (e.g. what species or group is of concern, what is the scope of the problem, what is the application need of the data (an in-lab support study or an in-field tool to determine probable cause, for instance), (2) the generation of reference databases built not necessarily with the newest technology, but using marker systems that are capable of answering the management or legal question and will be useful for some length of time (it will be cost prohibitive to rebuild a genetic database every few years with whatever new marker that comes along), (3) the archiving of reference samples (plant

tissue or DNA) that can be shared among labs, and (4) testing material in the usually short timeframe that meets law enforcement and management needs.

References

Durand, J, C Bodénès, E Chancerel, JM Frigerio, G Vendramin, F Sebastiani, A Buonamici, O Gailing, HP Koelewijn, F Villani, C Mattioni, M Cherubini, PG Goicoechea, A Herrán, Z Ikaran, C Cabané, S Ueno, F Alberto, PY Dumoulin, E Guichoux, A de Daruvar, A Kremer, and C Plomion. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. BMC Genomics 11:570. doi: 10.1186/1471-2164-11-570.

Eckert, AJ, B Pande, ES Ersoz, MH Wright, VK Rashbrook, CM Nicolet, and DB Neale. 2009. High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). Tree Genetics & Genomes 5:225–234.

Hodel, RG, MA Gitzendanner, CC Germain-Aubrey, X Liu, AA Crowl, M Sun, JB Landis, MC Segovia-Salcedo, NA Douglas, S Chen, DE Soltis, and PS Soltis. 2016. A new resource for the development of SSR markers: Millions of loci from a thousand plant transcriptomes. Application in Plant Sciences. 4(6). pii: apps.1600024. doi: 10.3732/apps.1600024.

Wu, B and T Hogetsu. 2009. Development and characterization of 11 microsatellite markers in *Lithocarpus edulis*. Conservation Genetics 10:1549–1551.

Chapter 2.4.

How New Generation Sequencing and genotyping technologies can help development of DNA traceability tools

Patricia Faivre Rampant[1], Berline Fomeju[1], Jean Noel Galliot[2], Aurélie Bérard[1], Aurélie Chauveau[1], Isabelle Le Clainche[1], Elodie Marquand[1], Dominique Brunel[1], Anne Farrugia[3], and Marie Christine Le Paslier[1]

[1] INRA-EPGV US 1279, France; [2] INRA, UREP 0874, France; [3] INRA, UMRH 1213, France

Background

Nowadays, thanks to the Next Generation Sequencing (NGS), genotyping technologies, and bioinformatics tools, it is possible to develop molecular resources to decipher new fields of research on orphan species or enhance program on ecological genomics. This can be done in an affordable cost, in a reasonable experimental and analysis time.

We reported here two cases to illustrate what it can be done from scratch with the help of NGS to identify species, launch genetic diversity studies on orphan species. Moreover, we demonstrated the accuracy of the multispecies SNP (Single Nucleotide Polymorphism) genotyping array as possible tool for traceability system.

Concomitant insect and plant metabarcoding

Semi-natural grasslands are considered as a vital habitat for wild pollinators, which consequently contribute to preserve the floristic diversity of this environment. Faced with the decline of pollinators, it is relevant to strengthen our understanding of the whole plant-pollinator interaction network. We demonstrated that the development of the NGS-DNA metabarcoding approach is powerful to answer the question: who visits what and who transports what? We developed a simple robust and quick workflow to identify at the same time the flower foraging insect and the pollen load included crude DNA extraction, targeted PCR amplification, NGS and public database queries (Figure 1). Flower-foraging insects were caught from the beginning of May to the end of July along three contrasted dairy farming systems in France (West, Center and Est). Sampling was carried out along six walking transects for each farming system. The results from more than 1000 flower visitor insects showed that the workflow is robust and easy to manage at low cost (Galliot et al, 2017).

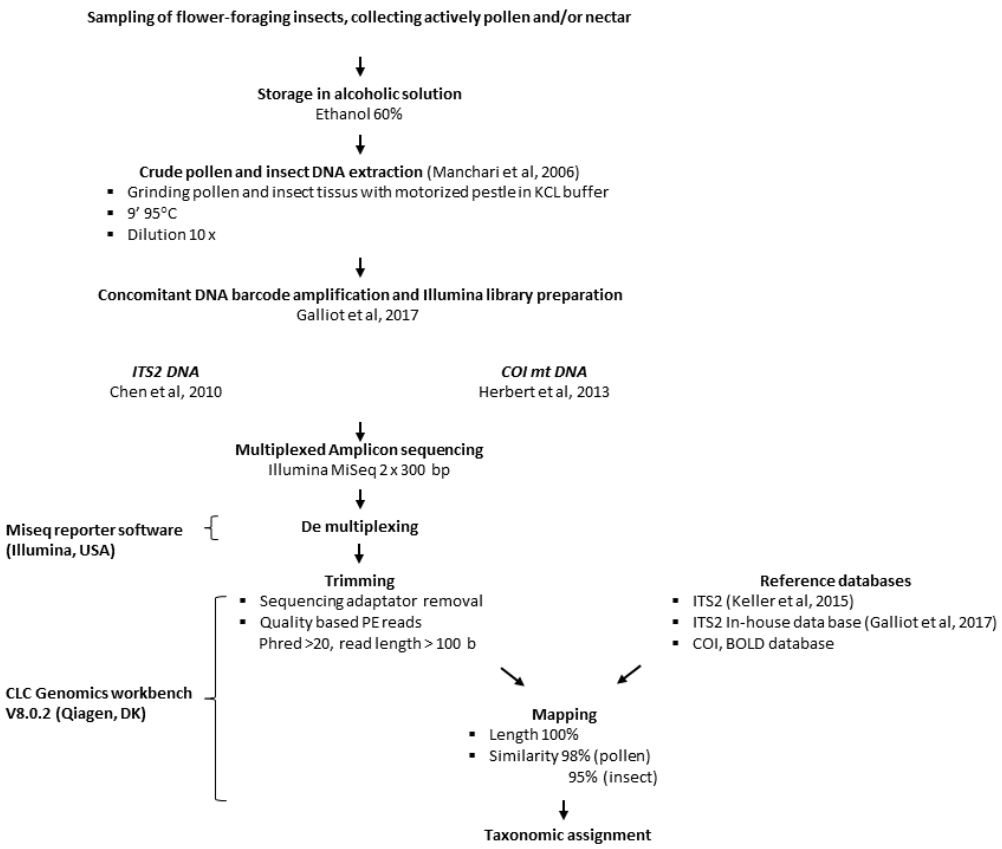


Figure 1. *Workfow of insect and plant metabarcoding. PE, Paired End reads; ITS2, internal transcribed spacer 2 region of nuclear ribosomal DNA; COI mt, mitochondrial gene cytochrome oxydase*

Creation of SNP collection in lavender, an orphan species

The lavender (*Lavandula angustifolia* Mill.) is a perennial sub-shrub plant that belongs to the Lamiaceae (or mint) family. Native to the Mediterranean region, this species has an economic importance in France, with a heritage and cultural significance. Moreover, it is cultivated worldwide for its essential oils. The lavender is one of the major crop of the Perfume, Aromatic and Medicinal Plants (PAM) sector and has many applications in industries such as pharmaceuticals, perfumes, cosmetics, and alternative medicine. Furthermore, the lavender cultivation contributes to the development of arid agricultural land in the south-east of France. For last decades, enhanced by global warming, lavender production is threatened by the Stolbur phytoplasma (*Candidatus Phytoplasma solani*) related disease. Additionally, in the PAM sector, the detection and the identification of helpful molecular markers is needed. These needs are mainly to study the genetic diversity of germplasms, to stay at the cutting edge of fraud prevention by certificating the varieties authenticity. In this context, we proposed a quick and low cost SNP detection procedure to develop further molecular tools, genetic diversity, QTL mapping and DNA traceability tools. We combined RNA-Seq data and DNA-Seq data to build a gene space of *Lavandula*. Firstly, a *de novo* transcriptome assembly was performed. Then, to recover the full-length gene sequences, we conducted an iterative

targeted DNA assembly using the iPEA protocol developed in our laboratory (Aluome *et al.* 2016). Finally, a collection of *Lavandula* cultivars was used to detect SNP using the gene space as reference. To our knowledge, this is the first study reporting SNP development for *Lavandula* (Fomeju et al, in preparation).

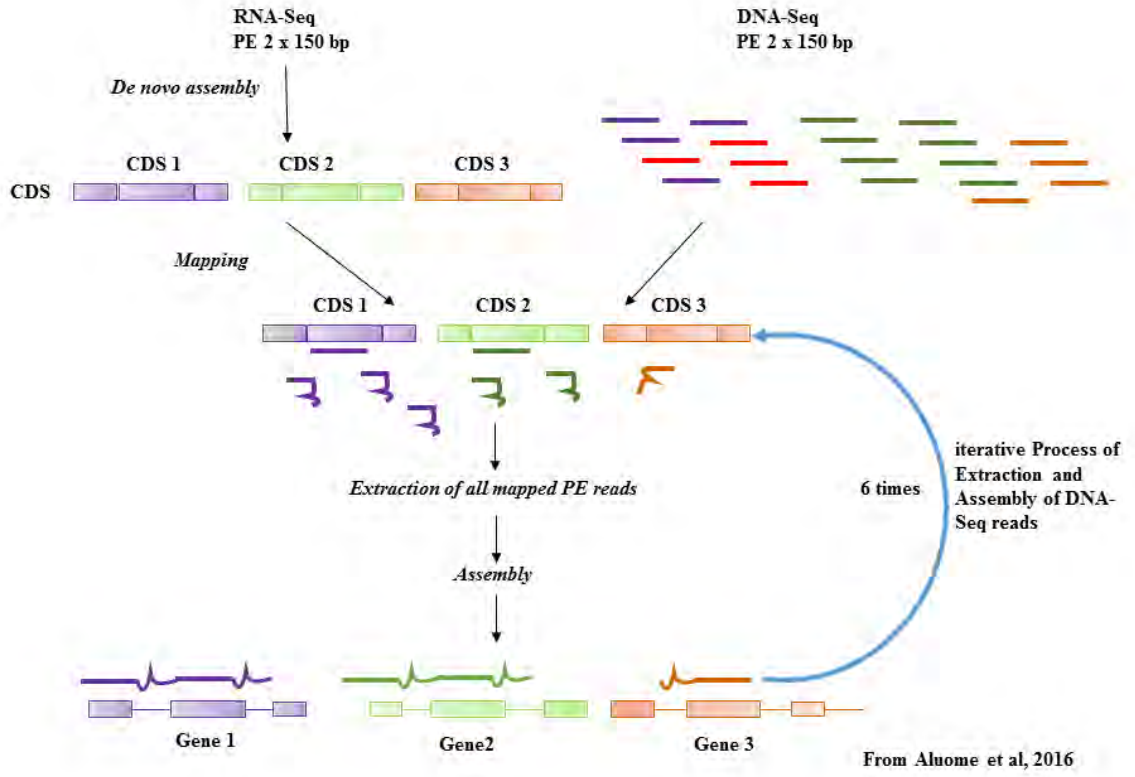


Figure 2. *Construction of the Lavandula gene space for SNP identification. CDS, coding sequences; PE, Paired End reads;* — *mapped reads;* ⌇ *unmapped reads*

Accuracy of a multispecies SNP genotyping array

Genotyping array provide data with good accuracy for thousands of SNPs (single nucleotide polymorphism) in thousands individuals for many organisms including plants. These extensive data used in large studies gave new insight in genetics. Moreover, genotyping experimental procedure could be performed quickly and genotyping calling do not required bioinformatics tools and skills. Nevertheless, flexible genotyping tools with affordable costs are still to be developed. Low cost genotyping assays would allow many applications including the control of individuals, the marker assisted management and genetic approaches in species where molecular developments have a comparatively low value relative to the cost of SNP arrays. The initial cost and/or the minimum sample number requirement are limiting factors for developing a new genotyping tool. An add-on option on already existing product gives a first opportunity to reduce the cost of a beadchip. Designing a multi-species array could be another alternative to reach the minimum sample number requirement while dropping prices per species. We successfully validate the genotyping accuracy of a multi-species-plant beadchip. Four custom Illumina bead pools, originally manufactured for pea (Tayeh et al, 2015), rapeseed (Chalhoub et al, 2014), grape (Le Paslier et al, 2016) and poplar (Faivre Rampant et al, 2016)

genomic studies developed at INRA (French National Institute of Agriculture), were then combined in a single array. Individuals/DNA already genotyped were newly genotyped with this 4-species array. For each SNP, the reproducibility in terms of genotype call rate, cluster separation was over than 0.9 (Le Paslier et al, 2016).

We demonstrated that NGS and genotyping tools can lead to the development of molecular tools starting from scratch for genetic studies, species identification at relatively low cost. NGS technologies offer many options for SNP discovery and genotyping; new protocols based on probe hybridization, extension and NGS, more flexible, accurate and cost effective are under deployment. Moreover, with any doubt, the forthcoming real time sequencing technology will bring new era for the development of quick, efficient, affordable molecular tools for DNA traceability of biological samples.

Acknowledgements: We acknowledge A. Boland-Augé group from CEA-Institut François Jacob-CNRGH, Evry for performing DNA samples QC, M.T Bihoreau and D. Lechner for providing INRA-EPGV group access to the Illumina Sequencing Platform.

References

Aluome *et al.* 2016. *De novo* construction of a "Gene-space" for diploid plant genome rich in repetitive sequences by an iterative Process of Extraction and Assembly of NGS reads (iPEA protocol) with limited computing resources. BMC Research Notes, 9, 81. doi:10.1186/s13104-016-1903-z

BOLD database http://www.boldsystems.org

Chalhoub B. *et al.* 2014. Plant genetics. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science, 345(6199): 950-95

Chen S. *et al.* 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PLoS ONE 5, e8613

Faivre-Rampant P. *et al.* 2016. New resources for genetic studies in *Populus nigra*: genome wide SNP discovery and development of a 12k Infinium array. Molecular Ecology Resour. 16: 1023–1036. doi:10.1111/1755-0998.12513

Galliot *et al.* 2017. Investigating a flower-insect forager network in a mountain grassland community using pollen DNA barcoding. J. Insect Conserv. DOI 10.1007/s10841-017-0022-z

Hebert P.D.N. et al, 2013. A DNA 'Barcode Blitz': Rapid Digitization and Sequencing of a Natural History Collection. PLoS ONE 8, e6853

Keller A. *et al.* 2015. Evaluating multiplexed next-generation sequencing as a method in palynology for mixed pollen samples. Plant Biol. 17(2):558–566

Le Paslier *et al.* 2016. Performance of a Multi-Species-Plant Illumina BeadChip, PAG XXIV - Plant and Animal Genome Conference, San Diego, USA. January 8-13 2016

Manchari *et al.* 2006. Weed response to herbicides: regional-scale distribution of herbicide resistance alleles in the grass weed *Alopecurus myosuroides*. New Phytologist, 171: 861–874. doi:10.1111/j.1469-8137.2006.01788.x

Le Paslier M. C. *et al.* The GrapeReSeq 18K *Vitis* genotyping chip. In Ninth International Symposium on Grapevine Physiology & Biotechnology; La Serena, Chili. April 21-26 2013

Tayeh N. *et al.* 2015. Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high density, high resolution consensus genetic map. Plant J. doi:10.1111/tpj.13070

Conclusions of PART 2

High-throughput technologies to identify tree variability

Search for biological information (including geographical distribution, reproductive systems, genetics, etc.) on available databases.

- - - - - - -

Build an international platform (GTTN2 would try to play this role) to integrate fragmented efforts from international research teams. These groups will accomplish the required analysis regarding uncovered prioritized species using reference samples.

- - - - - - -

Complex projects have to be designed to unravel genetic variation and, thus, discriminate the origin of given species.

- - - - - - -

Search for funding information to prepare collaborative actions to integrate efforts for creating resources.

- - - - - - -

Development of genotyping tools for traceability, based on molecular markers that allow species or species and origin discrimination:

- *First phase: Design multispecies genotyping technologies for cost-effective analysis.*
- *Second phase: Develop new strategies to perform analysis in mobile units for field-testing by producers, law enforcement bodies and officials.*

- - - - - - -

Need to promote the willingness practice by private sector of certification for wood-based products from their plantations.

PART 3

BIOINFORMATICS ANALYSIS AND DATABASE GENERATION BASED ON HIGH-THROUGHPUT INFORMATION

Chapter 3.1.

Cost-effective approaches for variant calling and analysis of complex plants

José De Vega
Crop Genomics Group, Earlham Institute, United Kingdom

Identifying genomic variation:

Without genetic variation, some of the basic mechanisms of evolutionary change cannot operate. Some variation is positive because it improves our ability to survive or adapt. The three primary sources of genetic variation are mutations, gene flow and sex. Within the first group, mutations, we can distinguish several classes, sorted by length:

- Whole-genome duplications (WGD).
- Chromosomal segmental mutations such as deletions, inversions, insertions and translocations (CSM).
- Structural variations (SV), which usually affect between 50 and 1000 bp.
- Indels, for short insertions and deletions (SNVs).
- But the most common form of genetic variants among individuals are the smallest, known as single nucleotide polymorphisms (SNPs).

Next-generation sequencing (NGS) can now sequence an entire human genome in a few days, and this capability has inspired a flood of new projects that are aimed at sequencing the genomes of thousands of individuals (Alexandrov *et al.* 2014). Current cost-effective NGS platforms produce shorter reads (Between 100 and 250 bp reads in the actual generation of Illumina technology) than Sanger sequencing, but with vastly greater numbers of reads.

When the polymorphisms are not known, discovery techniques are to analyse data blind to the possible location of the variation; stringent thresholds must therefore be applied to control false positives. By contrast, genotyping techniques offer increased power to detect polymorphisms once the variant is known, and more relaxed thresholds may be applied than for discovery (Alkan, Coe, and Eichler 2011).

Whole-genome resequencing combined with new, highly efficient alignment software is being used to discover large numbers of SNPs and structural variants in previously sequenced genomes. For SNP discovery, using raw reads can provide greater resolution than using a genome assembly. With raw reads, both the depth of coverage as well as the proportion of mixed alleles can be quantified, in contrast to creating an assembly, in which all coverage at a given locus is collapsed into a single base call. The basic workflow can be divided in two steps:

(a) aligning this great number of reads to the reference genomes, and (b) assess the differences in each position:

(a) To map reads to a reference genome, the location of each read, relative to the reference genome, is predicted. Many novel computational tools have been developed to map NGS reads to genomes and to reconstruct genomes and transcriptomes. Due to using different mapping techniques, each tool provides different trade-offs between speed and quality of the mapping (Hatem *et al.* 2013).

(b) Variant calling algorithms compare mapped reads to the reference genome and identify potential variants. SNP and indel calling algorithms vary in their approach to identifying candidate variants (Alkan, Coe, and Eichler 2011). Basic algorithms identify variants based the on the number of high confidence base calls that disagree with the reference bases for the genome position of interest. More sophisticated algorithms commonly use Bayesian, likelihood, or machine learning statistical methods (Pabinger *et al.* 2014). In the case of SNV, sequencing-based methods have used mate-pair or paired-end reads for structural variant discovery. In this approach, two paired reads are generated at an approximately known distance in the donor genome. Pairs mapping at a distance that is substantially different from the expected length, or with anomalous orientation, suggest structural variants (Medvedev, Stanciu, and Brudno 2009).

Sources of error in NGS-based variant calling:

Several types of errors can impact the accuracy of SNP and indel variant identification. These errors occur during (a) sample processing, (b) the chemical and electronic processes that occur during sequencing, as well as the (c) bioinformatic processing of sequence data: base calling, read mapping or de novo assembly, and identification of SNP and indel variants (Nielsen *et al.* 2011):

(a) Substitution and indel polymerase errors accumulate during amplification and then exponentially for subsequent cycles (PCR duplicates), at which point they may significantly contribute to variant call error (Kozarewa *et al.* 2009).

(b) A number of systematic sequencing errors have been described for the various sequencing platforms (Ross *et al.* 2013). Some systematic sequencing errors have distinct characteristics, such as strand bias, which can be used to distinguish them from likely true variant calls.

(c) In general, highly diverse regions of the genome are more prone to mapping and alignment errors than lower diversity regions (Nielsen *et al.* 2011). Reads that map to "duplicated" regions in the reference genome are usually given a low mapping quality score by the algorithm, which are typically filtered. Many variant callers often use additional filters such as a minimum depth of coverage threshold, base call frequency (e.g., > 90% of calls at a position being identical), masking of homopolymer and repetitive/duplicated sequence regions, and trimming of poor quality bases from the ends of reads. These filters can be hard filters with carefully chosen cut-offs, or can be chosen by machine learning algorithms (E.g. GATK's Variant Quality Score Recalibration).

Long-read sequencing platforms such as Oxford Nanopore technologies and Pacific Biosciences single-molecule real-time sequencing can achieve reads >10 kb. Observed error rates on both platforms are still high, but improving very quickly. While long-read sequencing is unlikely to be cost-effective for genotyping studies in the near future, improvements in existing reference genomes will benefit genotyping. As the costs of long-read sequencing decrease in the future, together with an increase in read quality, we can expect genotyping methods to make more use of this technology.

Coverage is key for a reliable SNP calling:

Sequence depth influences the accuracy by which rare events can be quantified in genomic quantification-based assays (Sims *et al.* 2014). Previous estimates of the amount of sequencing required to accurately identify SNPs in WGS and exome-seq are variable. Bentley *et al.* (2008) estimated that 15X mapped read depth of WGS samples would be sufficient to detect almost all homozygous SNPs and 33X for almost all heterozygous SNPs. 50X was estimated by Ajay *et al.* (2011) for all SNPs and small indels.

However, the ever-increasing demand for larger samples suggests that medium (5-20X) or low-coverage sequencing will be the most common and cost-effective study design in many applications of NGS for years to come. For example, the 1000 Genomes Project pilot (Consortium 2010) relied on approximately 3X coverage to sequence 176 individuals genome-wide. This design is more cost-efficient than deeper sequencing in fewer individuals. Likewise, in association studies, mapping power is typically maximized by sequencing many individuals at low depth, rather than sequencing fewer individuals at a high depth (Kim *et al.* 2010). In another example, Pasaniuc *et al.* (2012) show that extremely low-coverage sequencing (0.1-05X) captures almost as much of the common (>5 %) and low-frequency (1-5 %) variation across the genome, without an excess of false positives.

Approaches based on Genome reduced-representation (RRS):

A widely used range of methods for detecting SNPs using high-throughput sequencing are known as genotyping-by-sequencing (GBS), they share common bases. Since the inception of GBS, it has undergone continuous development, giving rise to at least 13 approaches (Scheben, Batley, and Edwards 2017)). It is important to notice that the Dutch company Keygene claims ownership of any RRS approach based on enzymatic digestion, and has successfully enforced it at least once in court (Keygene 2016).

In its more basic approach, GBS is a simple highly multiplexed system for constructing reduced representation libraries for the Illumina NGS platform developed in the Buckler lab (Elshire *et al.* 2011). It generates large numbers of SNPs for use in genetic analyses and genotyping (Beissinger *et al.* 2013). Key components of this system include low cost, reduced sample handling, fewer PCR and purification steps, no size fractionation, no reference sequence limits, efficient barcoding and easiness to scale up (Davey *et al.* 2011).

However, successful implementation of GBS in complex heterozygous crop is limited. Many reasons may have contributed to this lack of adoptions. From a technical perspective, missing

data, genotyping errors and heterozygote under-calling are common in GBS results due to uneven sequencing depth across sites and high level of sample multiplexing (Beissinger *et al.* 2013; Swarts *et al.* 2014).

Alternative approaches based on amplification instead of enzymatic digestion are AmpSeq (Yang *et al.* 2016) or rAmpSeq (Buckler *et al.* 2016), both are semi-automated pipelines based on amplicon sequencing that incorporates a machine learning model for primer design and uses Illumina's Nextera dual-barcoding and sequencing platforms for genotyping.

Alternative approaches based on hybridization using commercial bails, the most common implementation is exome-capture, but any know sequence can be targeted. The process of exome-seq has known issues that impact negatively on SNP detection sensitivity. These include PCR amplification, which tends towards lower coverage in GC-rich regions due to annealing during amplification, and the preferential capture of reference sequence alleles, which biases the allele distribution away from alternate alleles at heterozygous SNP sites. Exome-seq produces a relatively heterogeneous profile of read coverage over target regions when compared to the more homogeneous WGS. Better uniformity of coverage yields improved SNP detection sensitivity across the regions of interest. However, Exome-seq target capture technology is clearly improving in sensitivity, sensibility and cost (Meynert *et al.* 2014).

The amount of sequencing required to accurately identify SNPs in Genome reduced-representation are higher than WGS, and depends on the capture kit. Clark et al (2011) calculated that exome-seq required 80X mean on-target depth to reach the common threshold of 10X per-site depth in 90 % or more of all targeted regions. And Meynert et at (2013) on some of the original exome-seq target capture kits estimated between 20X and 46X mean on-target depth was required to successfully genotype 95 % of heterozygous SNPs.

Cost-effective low-coverage whole-genome sequencing:

Whole-genome resequencing (WGS) differs from RRS in the lack of complexity reduction steps before sequencing. In a WGR approach known as skim genotyping-by-sequencing (SkimGBS), SNPs and genotypes are called using low-coverage genomic reads, typically <1X, to make genotyping large populations viable (Bayer *et al.* 2017). Uniformity of coverage is clearly still a major issue for exome sequencing in terms of capturing a reasonable number of reads across all of the targeted regions. PCR amplification-free library preparation can mitigate the issue somewhat for WGS samples.

To simplify data analysis, heterozygous alleles are often eliminated by sequencing recombinant inbred line (RIL) or double-haploid (DH) populations (Scheben, Batley, and Edwards 2017). In Huang et al (2009) the recombinant lines were sequenced to an average coverage of 0.02X, identifying a total of 1,493,461 SNPs, with an average density of 1 SNP every 40 kbps. The parental genomes and a reference sequence are often required for these mapping populations (Golicz, Bayer, and Edwards 2015; Huang *et al.* 2009), although they can also be inferred using statistical models. Training the model on each individual sample refines this approach by allowing for variation in error rates (Scheben, Batley, and Edwards 2017).

Cost per marker is obviously significantly cheaper than RRS in whole-genome approaches. A rough estimate of the costs of WGR is <0.0001 USD per marker or approximately $80 per sample. However, the cost per sample will remain higher in WGR than RRS (Scheben, Batley,

and Edwards 2017). On the other hand, when sample preparation cost and imputation are taken into account, there is an optimal number of samples to sequence for any budget. Under zero sample preparation cost and ignoring the benefit of imputation, the optimal study design involves sequencing a maximal number of samples at minimal coverage (Kim *et al.* 2010). Because of this, a very cost-effective approach is to sequence DNA from pools of individuals, or Pool-seq (Anand *et al.* 2016). However, pooling of DNA creates new problems and complexity in data analysis. One of the most challenging problems of Pool-seq is to correctly identify rare variants (allele frequency, AF < 0.01), as sequencing errors confound with the alleles present at low frequencies in the pools. The power of many genetic analyses depends upon accurate determination of AFs of variants. In principle, Pool-seq give more robust estimate of AF due to the larger sample size (Anand *et al.* 2016). But sequencing noises might give rise to many spurious rare variants in Pool-seq and proper care should be taken to remove them before doing any kind of association studies (Rellstab *et al.* 2013; Anand *et al.* 2016).

My group at Earlham Institute in Norwich, UK, has tested in complex crops a cost-efficient whole-genome genotyping method based on low-coverage sequencing of individually-barcoded accessions. The method allows the identification of variants in individual samples at a cost equivalent to pool-seq. Following an iterative approach, outliers are re-pooled and sequenced to a higher coverage as required. The information of the different subpopulations is used to fine tune the variant calling. In summary, our approach provides an efficient and economical means of producing data for the design of high-density SNP genotyping platforms for species with draft sequence assemblies and provides a framework for methods in species that lack genome sequence.

Conclusion

The strategy of using next-generation sequencing has completely changed the landscape of high-density SNP assay development, particularly for non-human genome applications. It is no longer necessary to conduct distinct projects to identify putative SNPs and then validate and characterize their allele frequencies in target populations. It is apparent that it will not be feasible to shift all experiments to the analysis of sequences from separate individuals even with further reductions in sequencing costs. Thus, cost-effective sequencing methods will remain an important research tool for species with sufficiently large population sizes that permit the acquisition of adequate sample sizes. Genotype calling and SNP calling for NGS data have matured from simple methods based on counting alleles to sophisticated methods that provide probabilistic measures of uncertainty, and they can incorporate information from many individuals and linked sites. NGS will be central in genomic and genetic studies for years to come, and it is worthwhile now to focus attention on forming a solid foundation for future research in these areas.

References

Ajay, Subramanian S, Stephen CJ Parker, Hatice Ozel Abaan, Karin V Fuentes Fajardo, and Elliott H Margulies. 2011. 'Accurate and comprehensive sequencing of personal genomes', *Genome research*, 21:1498-505.

Alexandrov, Nickolai, Shuaishuai Tai, Wensheng Wang, Locedie Mansueto, Kevin Palis, Roven Rommel Fuentes, Victor Jun Ulat, Dmytro Chebotarov, Gengyun Zhang, and

Zhikang Li. 2014. 'SNP-Seek database of SNPs derived from 3000 rice genomes', *Nucleic acids research*, 43:D1023-D27.

Alkan, Can, Bradley P Coe, and Evan E Eichler. 2011. 'Genome structural variation discovery and genotyping', *Nature Reviews Genetics*, 12:363-76.

Anand, Santosh, Eleonora Mangano, Nadia Barizzone, Roberta Bordoni, Melissa Sorosina, Ferdinando Clarelli, Lucia Corrado, Filippo Martinelli Boneschi, Sandra D'Alfonso, and Gianluca De Bellis. 2016. 'Next Generation Sequencing of Pooled Samples: Guideline for Variants' Filtering', *Scientific reports*, 6.

Bayer, Philipp E, Bhavna Hurgobin, Agnieszka A Golicz, Chon-Kit Kenneth Chan, Yuxuan Yuan, HueyTyng Lee, Michael Renton, Jinling Meng, Ruiyuan Li, and Yan Long. 2017. 'Assembly and comparison of two closely related *Brassica napus* genomes', *Plant biotechnology journal*.

Beissinger, Timothy M, Candice N Hirsch, Rajandeep S Sekhon, Jillian M Foerster, James M Johnson, German Muttoni, Brieanne Vaillancourt, C Robin Buell, Shawn M Kaeppler, and Natalia de Leon. 2013. 'Marker density and read depth for genotyping populations using genotyping-by-sequencing', *Genetics*, 193:1073-81.

Bentley, David R, Shankar Balasubramanian, Harold P Swerdlow, Geoffrey P Smith, John Milton, Clive G Brown, Kevin P Hall, Dirk J Evers, Colin L Barnes, and Helen R Bignell. 2008. 'Accurate whole human genome sequencing using reversible terminator chemistry', *Nature*, 456:53-59.

Buckler, Edward S, Daniel C. Ilut, Xiaoyun Wang, Tobias Kretzschmar, Michael A. Gore, and Sharon E. Mitchell. 2016. 'rAmpSeq: Using repetitive sequences for robust genotyping', *bioRxiv*.

Clark, Michael J, Rui Chen, Hugo YK Lam, Konrad J Karczewski, Rong Chen, Ghia Euskirchen, Atul J Butte, and Michael Snyder. 2011. 'Performance comparison of exome DNA sequencing technologies', *Nature biotechnology*, 29:908-14.

Consortium, Genomes Project. 2010. 'A map of human genome variation from population-scale sequencing', *Nature*, 467:1061-73.

Davey, John W, Paul A Hohenlohe, Paul D Etter, Jason Q Boone, Julian M Catchen, and Mark L Blaxter. 2011. 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics*, 12:499-510.

Elshire, Robert J, Jeffrey C Glaubitz, Qi Sun, Jesse A Poland, Ken Kawamoto, Edward S Buckler, and Sharon E Mitchell. 2011. 'A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species', *PloS one*, 6:e19379.

Golicz, Agnieszka A, Philipp E Bayer, and David Edwards. 2015. 'Skim-based genotyping by sequencing', *Plant Genotyping: Methods and Protocols*: 257-70.

Hatem, Ayat, Doruk Bozdağ, Amanda E Toland, and Ümit V Çatalyürek. 2013. 'Benchmarking short sequence mapping tools', *BMC bioinformatics*, 14:184.

Huang, Xuehui, Qi Feng, Qian Qian, Qiang Zhao, Lu Wang, Ahong Wang, Jianping Guan, Danlin Fan, Qijun Weng, and Tao Huang. 2009. 'High-throughput genotyping by whole-genome resequencing', *Genome research*, 19:1068-76.

Keygene. 2016. Accessed 01/10/2017. (http://www.keygene.com/press-release-keygenes-sbg-patent-upheld-by-the-uspto-after-ex-parte-reexamination.

Kim, Su Yeon, Yingrui Li, Yiran Guo, Ruiqiang Li, Johan Holmkvist, Torben Hansen, Oluf Pedersen, Jun Wang, and Rasmus Nielsen. 2010. 'Design of association studies with pooled or un-pooled next-generation sequencing data', *Genetic epidemiology*, 34:479-91.

Kozarewa, Iwanka, Zemin Ning, Michael A Quail, Mandy J Sanders, Matthew Berriman, and Daniel J Turner. 2009. 'Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+ C)-biased genomes', *Nature methods*, 6:291-95.

Medvedev, Paul, Monica Stanciu, and Michael Brudno. 2009. 'Computational methods for discovering structural variation with next-generation sequencing', *Nature methods*, 6:S13-S20.

Meynert, Alison M, Morad Ansari, David R FitzPatrick, and Martin S Taylor. 2014. 'Variant detection sensitivity and biases in whole genome and exome sequencing', *BMC bioinformatics*, 15:247.

Meynert, Alison M, Louise S Bicknell, Matthew E Hurles, Andrew P Jackson, and Martin S Taylor. 2013. 'Quantifying single nucleotide variant detection sensitivity in exome sequencing', *BMC bioinformatics*, 14:195.

Nielsen, Rasmus, Joshua S Paul, Anders Albrechtsen, and Yun S Song. 2011. 'Genotype and SNP calling from next-generation sequencing data', *Nature Reviews Genetics*, 12:443-51.

Pabinger, Stephan, Andreas Dander, Maria Fischer, Rene Snajder, Michael Sperk, Mirjana Efremova, Birgit Krabichler, Michael R Speicher, Johannes Zschocke, and Zlatko Trajanoski. 2014. 'A survey of tools for variant analysis of next-generation genome sequencing data', *Briefings in bioinformatics*, 15:256-78.

Pasaniuc, Bogdan, Nadin Rohland, Paul J McLaren, Kiran Garimella, Noah Zaitlen, Heng Li, Namrata Gupta, Benjamin M Neale, Mark J Daly, and Pamela Sklar. 2012. 'Extremely low-coverage sequencing and imputation increases power for genome-wide association studies', *Nature genetics*, 44:631-35.

Rellstab, Christian, Stefan Zoller, Andrew Tedder, Felix Gugerli, and Martin C Fischer. 2013. 'Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species', *PloS one*, 8:e80422.

Ross, Michael G, Carsten Russ, Maura Costello, Andrew Hollinger, Niall J Lennon, Ryan Hegarty, Chad Nusbaum, and David B Jaffe. 2013. 'Characterizing and measuring bias in sequence data', *Genome biology*, 14:R51.

Scheben, Armin, Jacqueline Batley, and David Edwards. 2017. 'Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application', *Plant biotechnology journal*, 15:149-61.

Sims, David, Ian Sudbery, Nicholas E Ilott, Andreas Heger, and Chris P Ponting. 2014. 'Sequencing depth and coverage: key considerations in genomic analyses', *Nature Reviews Genetics*, 15:121-32.

Swarts, Kelly, Huihui Li, J Alberto Romero Navarro, Dong An, Maria Cinta Romay, Sarah Hearne, Charlotte Acharya, Jeffrey C Glaubitz, Sharon Mitchell, and Robert J Elshire. 2014. 'Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants', *The Plant Genome*, 7.

Yang, Shanshan, Jonathan Fresnedo-Ramírez, Minghui Wang, Linda Cote, Peter Schweitzer, Paola Barba, Elizabeth M Takacs, Matthew Clark, James Luby, and David C Manns. 2016. 'A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker-assisted selection in grapevine', *Horticulture research*, 3:16002.

Chapter 3.2.

Timber Tracking Initiatives - Computational Support with TreeGenes

Jill L. Wegrzyn
Department of Ecology and Evolutionary Biology, University of Connecticut, USA

The objective of the GTTN/EFI organizations as presented at the Application of high-throughput genotyping technologies for forest tree species identification and timber tracking meeting in Madrid (September 2017) was to identify both genetic and computational strategies to develop, store, and access relevant information for species of interest. The invitation to this meeting provided an opportunity for us to understand the scope of the problem and identify areas where our existing computational infrastructure could assist.

The TreeGenes database is best described as a web-based Clade Organism Database (COD) hosted in an open-source platform known as Tripal. Tripal combines the page creation utility provided with the Drupal content management system with the Generic Model Organism Database (GMOD) schema known as CHADO (Figure 1). The TreeGenes database has served the forest genomics community with a primary focus on genetics data, including genetic maps, marker data, transcriptomes, and full genomes. TreeGenes has recently expanded its offering to include phenotypic data as well as environmental metrics for studies association mapping and landscape genomics studies.

This recent focus was enabled through the collection of detailed metadata and original data from association mapping and landscape genomics studies at the time of publication. This effort has been facilitated in conjunction with peer-review journals to provide a permanent accession number that references this data. TreeGenes provides a customized workflow that directs submissions based on users' answers to specific questions regarding their sequencing/genotyping design, experimental design, and final analysis (Figure 2). There is an emphasis on the collection of georeferenced data for studies involving common gardens or landscape sampling. In cases where exact positions cannot be released, or are not available, approximate or obscured coordinates are stored in the database. The data from these studies is released following publication acceptance or permission from the researcher.

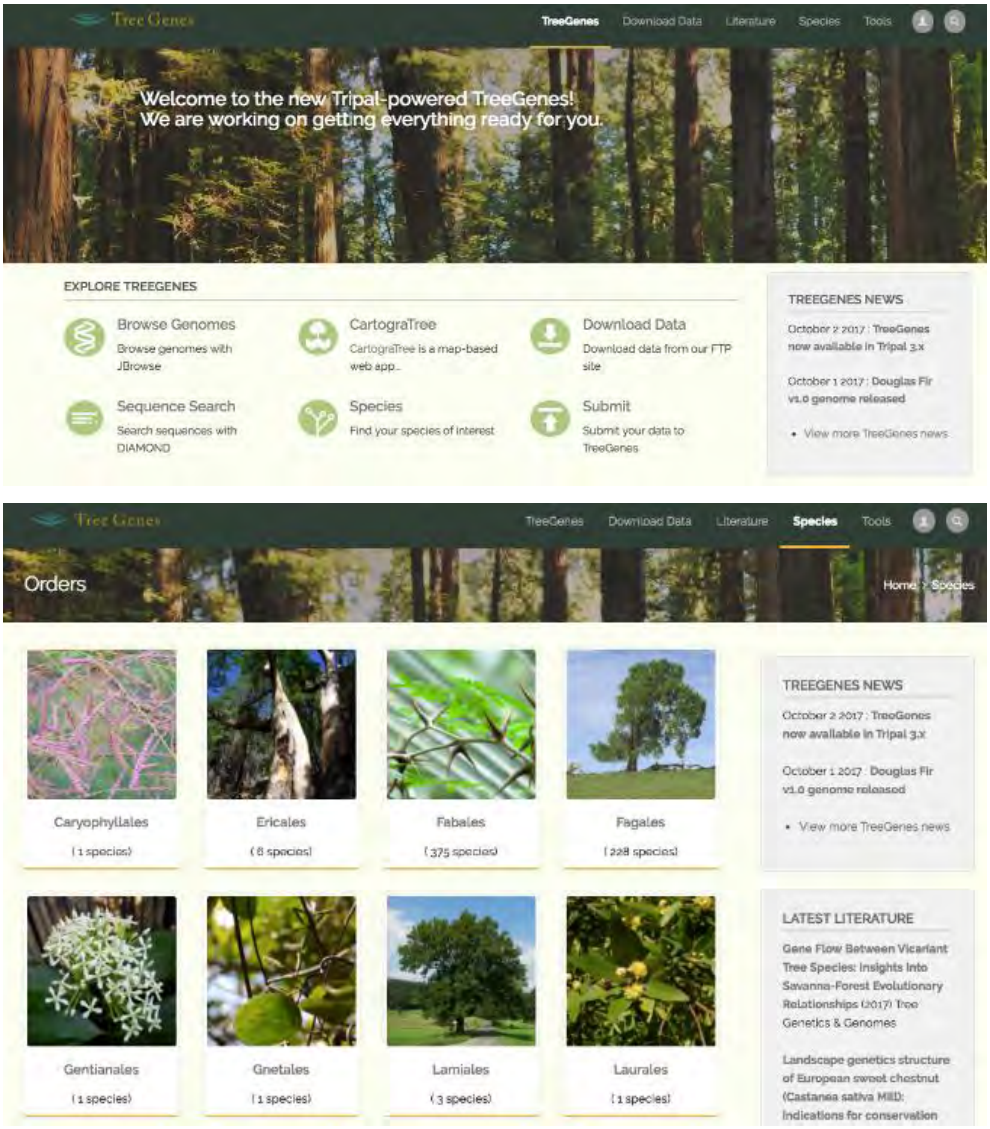


Figure 1. *TreeGenes Database*

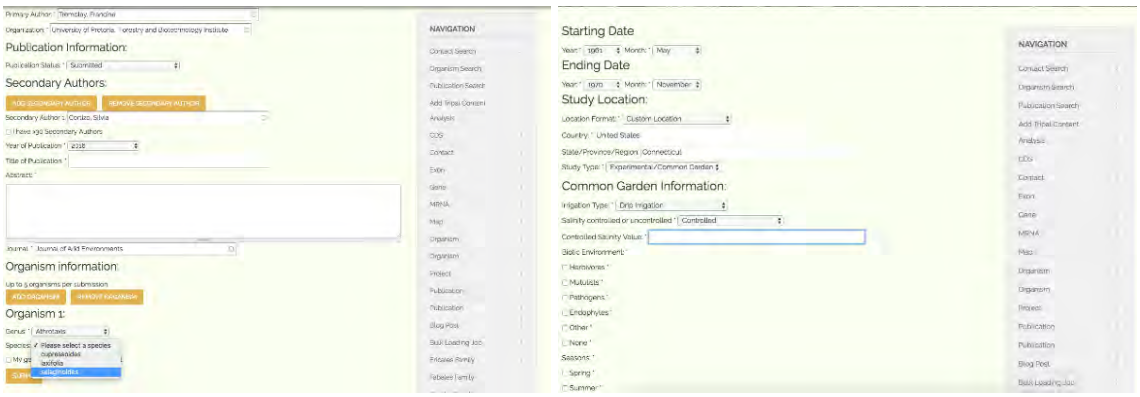


Figure 2:*.TreeGenes Data Repository Submission*

The data resulting from these collections is made available to the public in TreeGenes through an application known as CartograTree. This web-based platform integrates genotypic and phenotypic data alongside environmental layers to provide an integrative framework for visualization and analysis (Figure 3). This framework allows researchers to filter, query, and display data for single studies or across multiple studies. Current research is focused on developing more robust pre-filters so that data can be interrogated through dynamic visualization platforms, such as RShiny. CartogtraTree includes data analysis options through a connection to Tripal Galaxy which allows one to connect any next generation sequencing (NGS) dataset to analytical packages, such as short read alignment, SNP identification, and association mapping. Web services available in the latest release of Tripal (v3.0) streamline the ability to pick up datasets and connect them to available applications. The Galaxy component runs on either an application server attached to TreeGenes or through more substantial supercomputers. CartograTree allows researchers to visualize and analyze data associated with georeferenced trees without local computing resources.

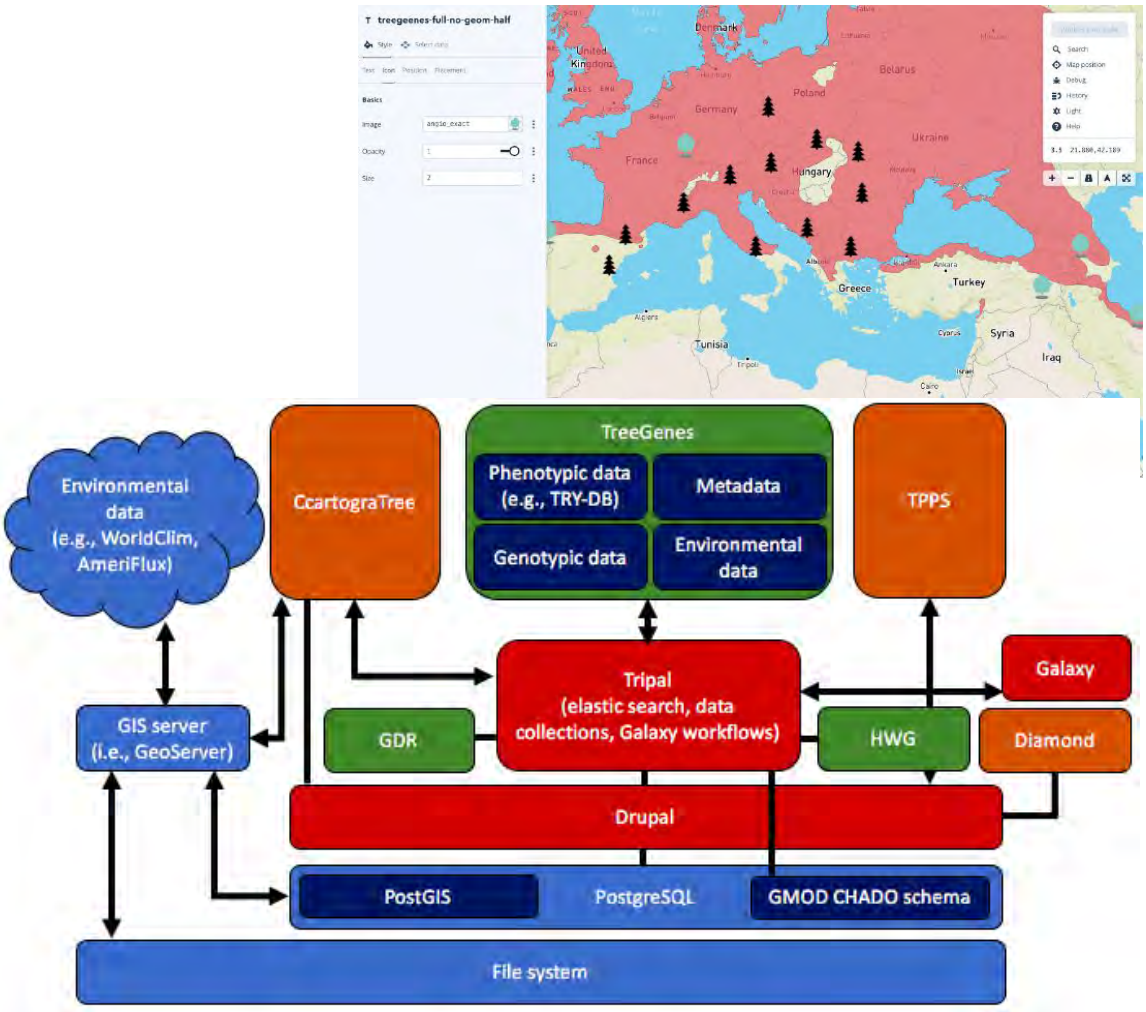


Figure 3: *CartograTree*

The adoption/modification of workflows in TreeGenes could help support timber tracking initiatives. In specific, the organizations involved have started to generate SNP and SSR data for a handful of populations. This data is often held with additional phenotypic information (ex: isotope data). GTTN is tasked with putting starting to database the species, the (human/lab) expertise, and the reference populations. Some of the primary concerns resulting from the workshop discussions include the following:

- Privacy
    - How much of the data identified from private labs should be made available publicly?
    - Should this data be displayed with true coordinates or not be displayed at all?
    - Can appropriate login control (group permissions) be used to provide individual labs with private space to work in?
- Data Integration
    - Should a central repository be available where information on the genotyped panel can be accessed by participating labs?
    - What types of data should be collected in addition to the marker data (SSR/SNPs)?
    - What efforts are underway to collect information on the species of interest?
    - Should internal records in the database also include location of wood sample and other data (physically) stored at various locations.
- Data Integrity
    - From a legal perspective, should minimum reporting standards be collected and enforced in a central location for reference samples?
    - What types of validation can be included at the time of data submission?
    - Should standard analytical pipelines be made available such that SSR or SNP data is subject to identical protocols?

From this discussion, TreeGenes provided information on the Tripal system which supports user logins and group level permissions to protect sensitive data and expose labeled data sets to specific users. In addition, TreeGenes has rapid development Drupal form templates that can be customized to the needs of individual labs. These forms support many types of validation that would be critical to ensure data integrity. TreeGenes has the ability to maintain information on the reference samples for each species but also keep the exact coordinates and/or metadata private to those parties involved in evaluating samples. For data that can be made public, it is easy to move datasets from the private space and into the public space for analysis with tools such as CartograTree. Finally, the Tripal Galaxy integration within TreeGenes allows us to configure specific analytical pipeline and exposure these to user groups. These pipelines could be initiated with the click of a button assuming the data is correctly formatted for analysis. This functionality could streamline consistent analysis for diagnostic labs.

Chapter 3.3.

Bioinformatics tools and resources to study plant genomes

Lieven Steck
Bioinformatics & Systems Biology, VIB, Belgium

OGA: a tool to analysing heterozygous crops

Although the advent of next generation sequencing (NGS) technologies offers new possibilities for characterising crop genomes, efficient SNP discovery tools suited to the analysis of allogamous species such as perennial ryegrass (*Lolium perenne*), with a complex genome and a high level of genetic heterogeneity, remains challenging. SNP discovery in crops can be based on transcriptome sequencing (RNA-seq) as an effective way to target exonic regions and avoid repetitive regions. *De novo* assembly algorithms relying on k-mer based De Bruijn Graphs (DBG) are appropriate for the assembly of large data sets of relatively short reads. These algorithms collapse near-identical reads into a single contiguous consensus sequence (contig). DBG *de novo* assembly algorithms typically tolerate low polymorphism density to discriminate between highly similar sequences such as duplicated regions, paralogues, or highly conserved domains. Consequently, divergent alleles are reconstructed as two independent contigs, despite representing the same locus. Furthermore, contig extension can be truncated in highly polymorphic regions, leading to neighboring, non-overlapping fragments of the same locus. As a consequence, *de novo* transcriptome assembly in highly heterozygous species typically yields a higher number of contigs than the actual number of genes expressed, thus rendering a redundant and fragmented reference transcriptome. However, a unique non-redundant consensus reference per expressed gene is essential for accurate read mapping and SNP discovery. Therefore, the assembly of a minimal redundant transcriptome in which allelic sequences have been collapsed while paralogues are kept separate is critical for accurate SNP calling. However, this is not a trivial task if the SNP discovery is to be conducted on RNA-seq data of multiple, genetically diverse genotypes that represent a broad range of the allelic diversity in the target germplasm.

Setting the sequence identity threshold for collapsing allelic sequences while keeping paralogues separate during *de novo* assembly, contig clustering, and annotation is complex. Plant genomes tend to contain a large number of multigene families, arising from tandem, segmental and whole genome duplication events. Paralogues may display differential

evolutionary divergence rates in various gene families depending on the age of the duplication event and evolutionary constraints. Likewise, the allelic variation is non-uniformly distributed across different functional classes of genes in the genome. Furthermore, polymorphisms are non-uniformly distributed across the length of the transcript due to domain-specific functional selection pressure. As a consequence, the degree of sequence similarity between paralogous sequences as well as between alleles, displays great variation and the ranges (percent sequence identity) probably partially overlap if considered across all gene families. In crops, this problem is exacerbated by many species being obligate cross-pollinators and some genotypes having high levels of allelic diversity. In line with general practice for transcript clustering, we assume that the sequence similarity between paralogous sequences is lower than that between allelic sequences, and that CAP3 collapses allelic sequences, while keeping paralogous sequences separate.

The strategy to get to a unique non-redudant consensus reference is demonstrated using an RNA-seq dataset of 14 unrelated *L. perenne* genotypes chosen as SNP discovery panel for an association mapping study. First *de novo* assembly using the CLCbio Genomics Workbench assembler (CLCbio) is performed on separate genotypes of L. perenne to minimise the level of sequence complexity (step 1). The number of contigs generated per genotype varied between 49,000 and 79,849, depending on the number of paired-end reads available per genotype. This resulted in a total of 930,208 contigs combined over all 14 genotypes. As a consequence, this creates a high degree of redundancy in the complete set of contigs (step 2) because commonly expressed genes are reconstructed in several genotypes. Comparison of the transcriptome composition of individual genotypes showed that 1,448 genes were assembled in only one genotype, 13,964 genes were assembled in at least five different genotypes, and 6,555 genes were assembled in all 14 genotypes. Each individual genotype contained on average about 70% of the total number of genes. Hence, the *de novo* assembled transcriptome of a single genotype cannot be used directly for read mapping because the transcriptome of any single genotype is incomplete in comparison to the total gene-space expressed of all genotypes. In addition, within any given genotype, allelic redundancy and transcript fragmentation is created for a subset of genes. Because this depends on the genotype-specific combination of alleles for any given gene, it is impossible to choose a single genotype in which all genes are assembled without allelic redundancy or fragmentation. In subsequent steps, transcript fragmentation and allelic redundancy thus need to be resolved. We use the predicted gene set of *Brachypodium distachyon* to guide further contig clustering and annotation. Therefore, we combine the primary DBG assembly with a subsequent Overlap-Layout-Consensus (OLC) assembly algorithm to generate a reference transcriptome with a single consensus reference sequence per locus. The procedure starts with a tBLASTn search of all *B. distachyon* proteins against the 930k *L. perenne* contigs. For each *B. distachyon* protein, all *L. perenne* contigs with a significant tBLASTn hit are grouped (step 3). Initial manual curation of the selected *L. perenne* contigs per *B. distachyon* candidate gene revealed that BLASTn must be performed at the protein level, and set to relatively low stringency to capture fragments with relatively low levels of sequence conservation between *B. distachyon* and *L. perenne* (e.g. 5’ and 3’ terminal sequences). However, this also allows paralogous sequences to be included in the contig selection due to the presence of conserved domains within protein families. CAP3 then generates one or more contigs from each of these groups (step 4). We assume that the sequence similarity between paralogous sequences is lower than that between allelic sequences, and that CAP3 collapses allelic and fragmented sequences into a consensus sequence, while keeping paralogous sequences separate. After extracting the contiguous

consensus sequence of each cluster (step 5), each CAP3 contig is compared to all 26,632 *B. distachyon* proteins by BLASTx (step 6). If the best BLASTx hit is obtained with the *B. distachyon* protein that founded its CAP3 cluster, the CAP3 contig is retained and is named according to the 'founding' *B. distachyon* protein to denote putative orthologous pairs. A CAP3 contig with a best BLASTx hit to any other protein than the one founding its clustering step is discarded because it probably corresponds to a paralogous sequence. So, while paralogous sequences may be included by tBLASTn selection of contigs in step 3, CAP3 separates paralogues by sequence alignment, and BLASTx filtering removes the paralogous 'by-products' in step 6. A given *L. perenne* transcript may be reconstructed several times, once with its *B. distachyon* orthologue as 'founding' protein (most optimal form), and perhaps as paralogous 'by-products' in the assembly cycle of a close family member (sub-optimal form). Therefore, the BLASTx selection routine (step 6) is included to remove all redundancy except the orthologous *B. distachyon* - *L. perenne* transcript pairs. This allows the discrimination of alleles and paralogous sequences, and facilitates annotation of the selected contigs according to their most likely orthologue in *B. distachyon* (step 7). Next, for each contig, the CDS is determined as the longest ORF in the reading frame indicated by the BLASTx (step 8). Because most allelic divergence is expected in the 5'UTR and 3'UTR, a second round of CAP3 clustering is performed using only the CDS sequences to further reduce the remaining allelic redundancy, followed by a final round of CDS detection and protein translation (step 9). We named this procedure Orthology Guided Assembly.

Throughout the procedure, we evaluate the quality of the transcriptome assembly by analysing the degree and origin of redundancy, the total number of resulting contigs, their length distribution, and the number of unique genes of the model species for which orthologous transcripts were assembled (gene space coverage). Furthermore, CAP3 clustering creates a majority vote consensus sequence for each locus, which represents the most common base at each position within the dataset itself. Using 380.292 SNPs with a read depth >8 in each genotype, we estimated that in 95.9% of the cases the reference sequence was identical to the most frequently observed base at that position, confirming that we indeed created a reference according to the most common sequence in the 14 genotypes. Because the reference is a consensus, the number of reference-read mismatches is more equally distributed across genotypes, as compared to choosing an individual genotype from within the dataset as reference, or any external reference genotype. Hence, our approach reduces bias from read mapping stringency parameters and consequently improves the quality of read count and SNP frequency estimates. Finally, we identify a transcriptome-wide set of more than one million transcript-anchored SNP markers in *L. perenne*.

An additional advantage from this procedure is that we can establish valuable linking between transcript assemblies and associated SNP markers from different but related species. Applying the same approach on two other species, namely *Festuca* and *Phleum* yielded similar efficiency concerning the amount of non-redundant transcripts being assembled. Now, since the assemblies of these two additional species are 'anchored' to the same *Brachipodium* gene we can easily group the assembled transcripts and align them to each other as well as to the used *Brachipodium* anchor sequence. This results in a multiple sequence alignment allowing us to identify inter- and intra-specific makers between species as well as between different genotypes within a species.

PLAZA: a resource to analysing plant genomes

Flowering plants contain many genes, a majority of which have been created during the last 150 million years through small- and large-scale duplications. Gene duplication and retention in plants has been extensive and gene families are generally larger in plants than in animals. Furthermore, most, if not all plants have experienced one, but probably more, whole genome duplications in their evolutionary past.

As a means to study genome organization and evolution, several tools and software have been developed to discover genomic homology based on gene collinearity within and among species. Collinearity information can be applied to analyse segmental and WGD events, whereas cross-species genome conservation facilitates the analysis of chromosomal rearrangements, such as inversions, chromosomal fissions/fusions, and translocations. These and other modes of evolution form the basis of the field of comparative genomics in which we analyse the duplication history as well as evolutionary relationships between species.

A key challenge in comparative genomics is the reliable grouping of homologous genes (derived from a common ancestor) and orthologous genes (homologs separated by a speciation event) into gene families. Orthology is generally considered a good proxy to identify genes performing a similar function in different species. Consequently, orthologs are frequently used as a mean to transfer functional information from well-studied model systems, such as *Arabidopsis thaliana* or *Oryza sativa* (rice), to non-model organisms. In plants, utilization of orthology is not trivial, due to a wealth of paralogs (homologous genes created through a duplication event) in almost all plant lineages. Ancient duplication events preceding speciation lead to outparalogs, which are frequently considered as subtypes within large gene families. In contrast to this are inparalogs, genes that originated through duplication events occurring after a speciation event. Besides continuous duplication events (for instance via tandem duplication), many plant paralogs are remnants of whole genome duplications (WGDs). In flowering plants, the frequent WGDs in several lineages result in the establishment of one-to-many and many-to-many orthologs (or co-orthologs).

PLAZA, an online resource for plant genomics, had been developed to integrate and distribute comparative genomics data for both computational and experimental plant biologists. The first release, based on nine sequenced plant genomes, included various tools to easily retrieve specific data types, such as gene families, multiple sequence alignments, phylogenetic trees, and genomic homology. To accommodate the evolutionary analysis of an increasing number of available plant genomes, more powerful and streamlined computational pipelines were required as well as new tools to visualize genome information from multiple species. The latest version of PLAZA, includes a major update of the comparative genomics platform, which currently hosts fifty-two species together with a variety of new tools to browse gene families, study functional clustering, and explore multispecies collinearity data. In addition to the development of a new tool to identify complex gene orthology relationships, several methods for finding orthologs between two or more species have been implemented and offered to the users, each with its own strengths and weaknesses. Whereas Reciprocal best BLAST-Hit (RBH) detection between closely related species provides a practical solution to identify orthologs, it cannot deal with complex one-to-many or many-to-many orthologous relationships between more distantly related species. Although the construction of phylogenetic trees should offer the highest confidence to identify speciation events in gene

family trees, it has a relatively low gene coverage compared to sequence-based clustering methods, as trees could not be generated for all gene families. In PLAZA, 76,651 phylogenetic trees were constructed covering 81% of all protein-coding genes assigned to gene families. Besides heavy computational requirements, the method is also hampered by its sensitivity to differences in the topology of the gene tree compared to the species tree, which are used for reconciliation.

To detect orthologous gene relationships in plants with an enhanced robustness, an integrative approach was developed to identify orthologs on a gene-by-gene basis: the PLAZA integrative Orthology Viewer. The developed ensemble approach consists of four distinct orthology prediction methods: orthologous gene families inferred through sequence-based clustering with OrthoMCL, reconciled phylogenetic trees, colinearity information and multispecies Best-Hits-and-Inparalogs (BHI) families. The latter are based on the best BLAST hit for each species, extended with the inparalogous genes in each species. The integration of gene collinearity facilitates the detection of positional orthologs, namely genes with conserved genome organization between species. The combination of different methods for orthology detection, as implemented in the PLAZA platform, allows for the more accurate selection of orthologs. The Integrative Orthology Viewer displays for a query gene and its predicted inparalogs the associated orthologs, including the support from the different orthology methods. In addition, all links are provided to explore the supporting evidence and specific details of the individual predictions. For instance, the phylogenetic trees that served as the primary data source for the tree-based orthologs can be viewed and the user can evaluate the support of a specific speciation node.

The PLAZA platform is a user-friendly platform for small- and large-scale comparative sequence analyses of plant genomes. The latest version includes many new genomes and implements new methods for colinearity and orthology detection as well as improved graphics.

Chapter 3.4.

Developing an expert database and a reference database with the Global Timber Tracking Network

Tommi Suominen and ***Meaghan Parker-Forney***
European Forest Institute, Finland; World Resources Institute, USA

1. Introduction

Unsustainable and illegal logging as well as related trade cause many economic and ecological problems both in producer and in consumer countries. It is one of the major causes for the loss of biodiversity. Although instruments against such unsustainable and illegal practices have been established, there is a lack of understanding how this instruments function in practice and practical control mechanisms to identify the origin of timber and wood products. Such methods are one of the fundamental prerequisites for improving forest and wood product sustainability through efficient import controls or corresponding origin testing by industry and the trade.

Existing timber tracking systems use paper-based documentation of timber origin and use at all levels of the processing process. Alternative methods are more and more used – especially DNA fingerprints and stable isotopes. In certain cases, the combination of both methods, DNA-fingerprints and stable isotopes, has the advantage that a higher spatial resolution and stronger statistical power for the control system can be expected. In addition, there are also complementary technologies (e.g. visual and chemistry) which are also very relevant. The methods and their applications for timber tracking advanced a lot during recent years and continue to advance. The progress has been regularly discussed in international workshops and during conferences of the first phase of GTTN 1, coordinated by Bioversity International. GTTN phase 2, coordinated by the European Forest Institute (EFI), is largely based and built on the results and experiences of the first phase.

The main objective of GTTN will remain the promotion of the integrated use of innovative technologies such as DNA fingerprinting and stable isotope analysis together with other existing technologies to combat illegal logging and associated trade worldwide. Building on the results and experiences of the first phase, the second phase will continue to further develop and expand the network, seek new partnerships, new funding sources and collaboration.

The GTTN information services are one of the main outputs to be delivered from the GTTN phase 2. The service under development is divided by their target groups into two: a Service Providers Directory (SPD) and a Reference Data Service (RDS). The Service Providers Directory needs to provide users along the supply chain with information on species identification and

the verification of the origin (country, region, concession) of wood and wood products. Furthermore, it should provide information on available scientific methods and tools (genetics, stable isotopes and wood anatomy), and contact information on certified labs for sample analysis. The Reference Data Service needs to provide internal users (that have put their data into the database, signed a data sharing agreement, successfully participated in ring tests for standardisation and are ready to provide the lab services) with safe password protected access to a data repository on geographically annotated reference data (e.g. genetics & isotopes) for identified priority species. The service is a crucially needed to share the world's presently collected data for the application of the identification methods by industry, academia and regulatory authorities.

In the first phase of GTTN, the conceptual development of the GTTN service design went hand in hand with the software development. This resulted in three different iterations of prototype designs of the system, each building on the user feedback to the previous prototype. At that point, there was no separation of services, but both the SPD and RDS were developed as one service. In the GTTN phase one project one of the complicated things was agreeing on the reference data sharing between data suppliers, who are also the users of that data. An agreement on the data sharing agreement was reached only in the end of the project. This prevented the system to be populated with reference data during the project, and as the reference data and expert directory services were designed as one, this consequently prevented also the deployment of the expert database.

2. The Service Providers Directory

The GTTN 1 phase user interface was more detailed and utilised an approach that was based on the user making a “claim” – i.e. a shipment claims to be of a specific tree species or trade name and/or claiming to be coming from a specific country. The user fills information relevant to the claim, which is then utilised to identify suitable laboratories/people for analysing this claim, grouped by identification technology type (see figure 1).

After the GTTN phase 1 ended, the World Resources Institute (WRI) had the last GTTN prototype simplified, resulting in a simplified version that removed the reference data related functionality, and concentrated only on the SDP. This result was still re-implemented by WRI as the SPD in a Drupal content management system (CMS) web service under the WRI Forest Legality Initiative website http://www.forestlegality.org.

The WRI SPD has essentially simplified the data structure to a list that can be filtered by different criteria, operated from four main level dialog boxes. The SPD version described here has not been published to general use. This makes the graphical user interface very simple (see figure 2). The design is geared to a user who knows the claimed species or a common name and is able to make the choice of a suitable analysis technology.

Now, EFI and WRI have teamed up to finalise the implementation of the Service Providers Directory, in a team effort. The finalisation of the expert database needs to be a top priority. There is a real and urgent need for enforcement authorities to know where they can get analysis competence from.

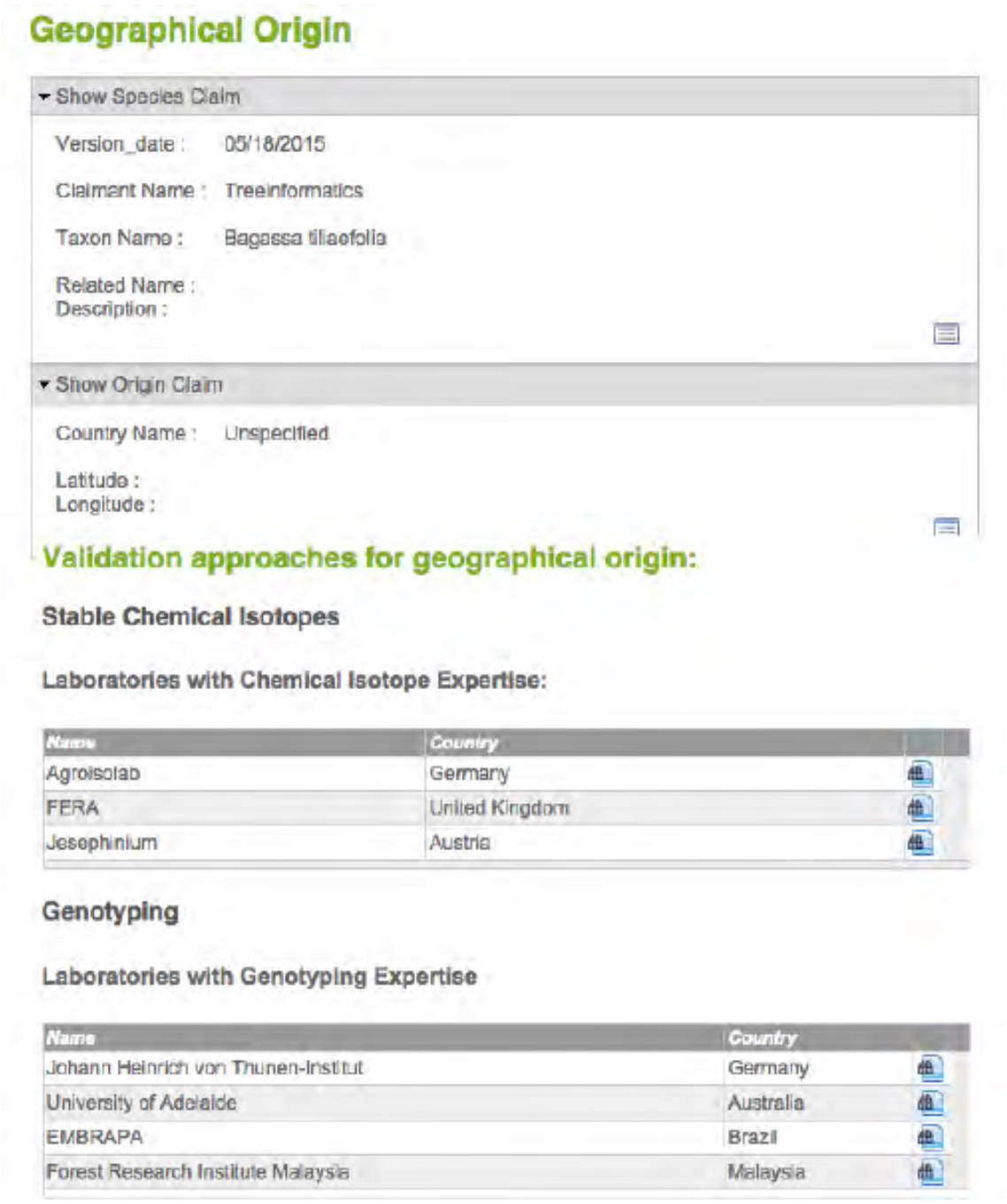


Figure 2 *- GTTN phase 1 claim-based user interface (source: GTTN Client User manual, 2015).*

The information on available expertise contained in the SPD is largely coming from an inventory conducted by TRAFFIC, but has been appended with information provided by the International Association of Wood Anatomists (IAWA). To ensure reliability and credibility of the SPD, this dataset that is 2-3 years old, needs to be made up to date. When external users start to use the SPD, first impressions are important for the willingness of the users to continue using the service and return to it. Out of date contacts will not be useful to any user, and would disincentivise them to return to the service. This would work against our objective to establish the SPD as a useful platform for support to e.g. enforcement authorities.

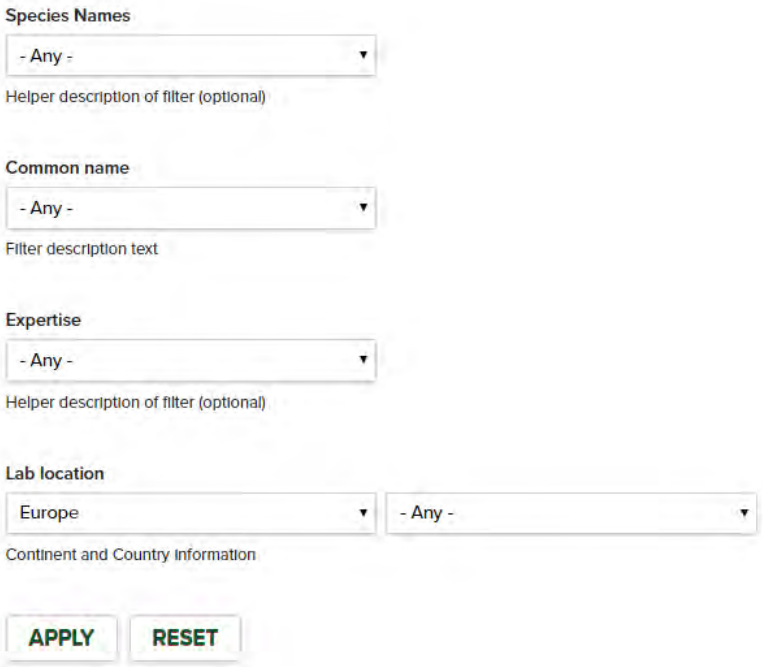


***Figure 3**. WRI simplified Service Providers Directory user interface.*

Before releasing the service, we plan to remove the contact details of individuals and implement an "email to service provider" functionality that allows contacting the experts with a service request, or a request for a quote for service. For the service providers, this will make it apparent that the work requests to do analysis work is facilitated through the SPD. We think this will on one hand incentivise the service providers to maintain the information on the SPD regarding what analysis competence they have, in order to avoid unnecessary contact requests or missing out on requests, if their competence description is too narrow. We will also need to make a facility to allow for the users to update their own contact and competence description data.

While we envision also more new features to improve the usability of this service, we will aim to release the service soon, as we understand the persistently reiterated need to get this service publicly available. We will carry out further user interface improvements to the service, after its initial release.

3. The GTTN Reference Data Service

The Reference Data Service is still in its design stage. Some of this functionality was also designed and implemented in the GTTN phase 1, but with this part we look at the design with fresh eyes, and try to analyse the needs and service design alternatives, before embarking on the implementation.

One of the main design choices under discussion currently is whether to build up a distributed system or a centralised system. Considerations for the two alternative design strategies are listed below.

A centralised system:

- Labs deposit directly their data in the database.
- Better for ensuring security, all access and data can be controlled.

A distributed system:

- Could use information from external, already existing data sources. The level of understanding that the GTTN database can have on what data these databases contain, depends on the level of metadata descriptions in those services.
- If a meta data service for this data is provided and the records/pages are described by metadata systematically, the situation is good.
- Inaccurate or only database level metadata means that we can only have a cursory understanding of the contents → integration potential weak.
- Allows to build on other organisations know-how, while they remain responsible for maintaining their own data
- Risk: can a distributed service be safe from external manipulation? If our service relies on external services, perpetrators could compromise the GTTN service by polluting the contents of any external service we are relying on for reference data.

A compromise between these two solutions could be a hybrid that contains both metadata describing the data available from participating labs by request and also contain actual data with its relevant metadata descriptions. Sharing of metadata implies no direct outbound linkage to external data, but producing an awareness of its presence, if labs are not inclined to share their data openly with everyone. A “data search” interface will lead the to the record describing the data and instead of a “Download data” button, there will be a button to “Send request to lab for reference data”.

Conclusions of PART 3

Bioinformatics analysis and database generation based on high-throughput information

Currently, GTTN is a meta-data-based, which could be re-structured and enriched in the future.

- - - - - - -

Development of integrative initiatives among research teams (including experts in taxonomy, genomics –sequencing, genotyping, etc-, bioinformatics, etc) would allow construction of an open-access database including raw data and shared information on reference samples.

- - - - - - -

The above-mentioned initiatives should firstly catalogue, evaluate and assess existing data resources (eg. NCBI, TreeGenes,…) and establish the key service and data providers in terms of reference materials.

- - - - - - -

The availability of such an open-resource requires proper funding for integration, curation, improvements, keeping it up-to-date and development of tools to facilitate data access to the final users/goals and database cross-talking. Funding strategy is therefore required at national and international level.

PART 4

APPLICATION OF DNA TECHNOLOGIES TO PREVENT TIMBER ILLEGAL LOGGING AND TRADING

Chapter 4.1.

Development and application of genetic reference data based on SNPs for timber tracking of tropical tree species

Bernd Degen and ***Céline Blanc-Jolivet***
Thünen Institut für Forstgenetik, Germany

Background

Since 2012, we are generating in frame of two large international projects genetic reference data based on SNPs for several tropical tree species. From 01/02/2012 to 31/07/2015, we worked together with 16 partners on the project "Development and implementation of a species identification and timber tracking system in Africa with DNA fingerprints and stable isotopes"(Degen and Bouda 2015). This project was funded by the International Tropical Timber Organization (ITTO) through grants from Germany, USA and Australia. Here, we focused on the economic important tree species Iroko (*Milicia excelsa*), Ayous (*Triplochiton scleroxylon*) and Sapelli (*Entandrophragma cylindricum*). The project collected more than 5400 leaf, cambium and wood samples as reference material over the distribution areas of the three species. The samples were taken from seven African countries: Cameroon, the Republic of the Congo, the Democratic Republic of the Congo, Ivory Coast, Gabon, Ghana and Kenya. To help capacity building and transfer technology, the project established three reference laboratories in tropical Africa: at the Forest Research Institute of Ghana in Kumasi for West Africa; at the Institut de Recherche en Ecologie Tropicale in Libreville, Gabon, for Central Africa; and at the Kenya Forestry Research Institute in Nairobi for East Africa. In October 2014, we started as a follow up with the "Large scale project on genetic timber verification". This project is financed by the German Federal Ministry of Agriculture and Food (BMEL) and aims to develop genetic reference data for timber tracking of additional seven tree species in Africa and seven species in Latin-America (https://www.thuenen.de/en/fg/projects/current-projects/largescale/). In Africa, samples were taken in Liberia, Ivory Coast, Ghana, Nigeria, Cameroon, Gabon and the two Congos. In Latin-America, collections of reference material have been done in Bolivia, Brazil, French Guiana and Peru.

Strategy for SNP development

In the last years, we have been focusing on timber tracking on SNPs (Single Nucleotide Polymorphism) because the fragment amplified are short and thus better suitable for degraded DNA of wood. Based on the experiences in former projects and the day to day work

experience of the Thünen-Centre of Competence on Timber Origin, we did the SNP development in 4 steps (figure 1):

1. Using next generation DNA sequencing techniques, we identified for each species more than 1000 SNPs. For this, we gained good experiences with the Rad-Sequencing (Pujolar *et al.* 2014) and "Skimming" (Besnard *et al.* 2014). The RAD-Sequencing is done with high coverage for a few individuals in order to obtain candidate SNPs in the nucleus. The "Skimming" is a Miseq-sequencing approach with low coverage that provides candidate SNPs in the plastidial genomes. The next step was the selection of a set of 480 SNPs by a pre-screening of several hundred SNPs at 96 individuals using the MassARRAY® iPLEX™ platform (McKernan *et al.* 2002).
2. Then we selected a set of 120 SNPs among the most informative ones for the final genetic screening of all 800-1500 individuals per species. The result of this step was the genetic reference data base.
3. The last step was aiming to simplify the work of timber testing by development of simple methods for the screening of the top 15 highly informative (selection of 15 "golden" or "silver" gene markers). Here we call a SNP a "golden" marker if it is fixed in one species or source of origin to one allele and for other species or sources of origin to another allele. For "silver" SNPs the fixation is not complete, but there are large differences at the allele frequencies among groups. Then, the highly informative SNPs were grouped in a single SNaPshot set for cost-effective screening.

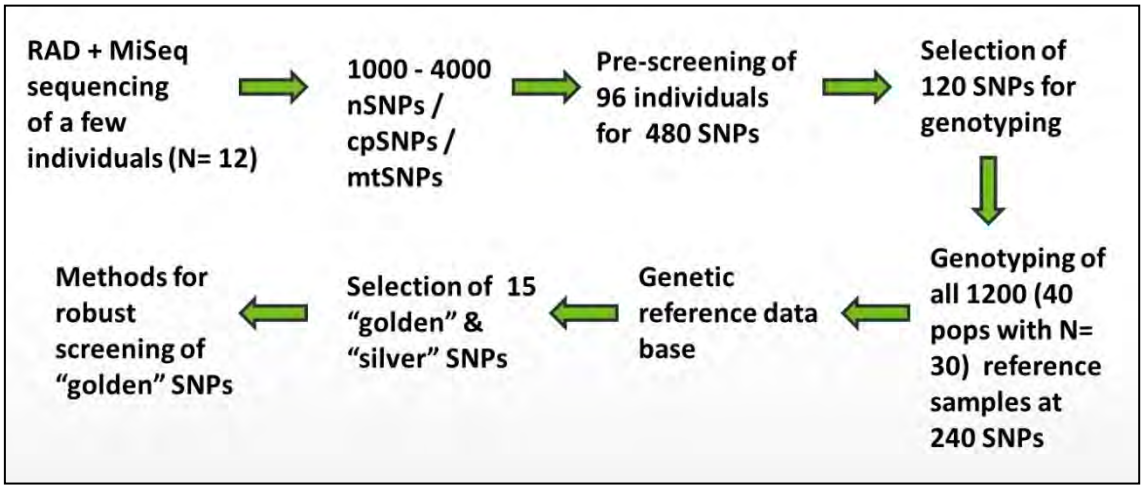


***Figure 1**. Strategy for the development of SNPs for timber tracking at the Thünen Institute of Forest Genetics*

As criteria for the selection of the final set of 120 SNPs (step 3), we were considering the general amplification rate for the SNPs, genetic differentiation among species and / or regions, and the correlation among genetic and spatial distances among regions and the type of information provided by each locus. Self-genetic assignment tests on the selected loci were compared with results from the complete set of loci to check that no geographical information is lost.

So far we have published the development strategy and the resulting SNPs for 4 species or genera from the above projects: Ayous (Jardine *et al.* 2016), Sapelli (Blanc-Jolivet *et al.* 2017a), Iroko (Blanc-Jolivet *et al.* 2017b), Khaya (Pakull *et al.* 2016). For Ayous, Iroko and Sapelli, we had not included the MiSeq approach, but this was the case for Khaya.

Examples

Iroko *(Milicia excelsa)*

In co-operation with partners in Belgium (University Brussels, Nature+) and France (INRA), we developed genetic reference data for Iroko based on a set of 55 SNPs screened at 1480 individual Iroko trees in eight African countries (Dainou *et al.* 2016, Blanc-Jolivet *et al.* 2017b, Dainou *et al.* 2017). There is a clear geographic pattern of genetic groups within the distribution range (figure 2). Western Africa can be separated from Central and East Africa. The genetic differentiation among countries measured with delta (Gregorius 1987) was with 0.172 moderate, same as the population fixation F_{ST} = 0.20. But all genetic groups were present in more than one country, and thus a self-assignment test gave only moderate success rates from 45% to 88% on the country level.

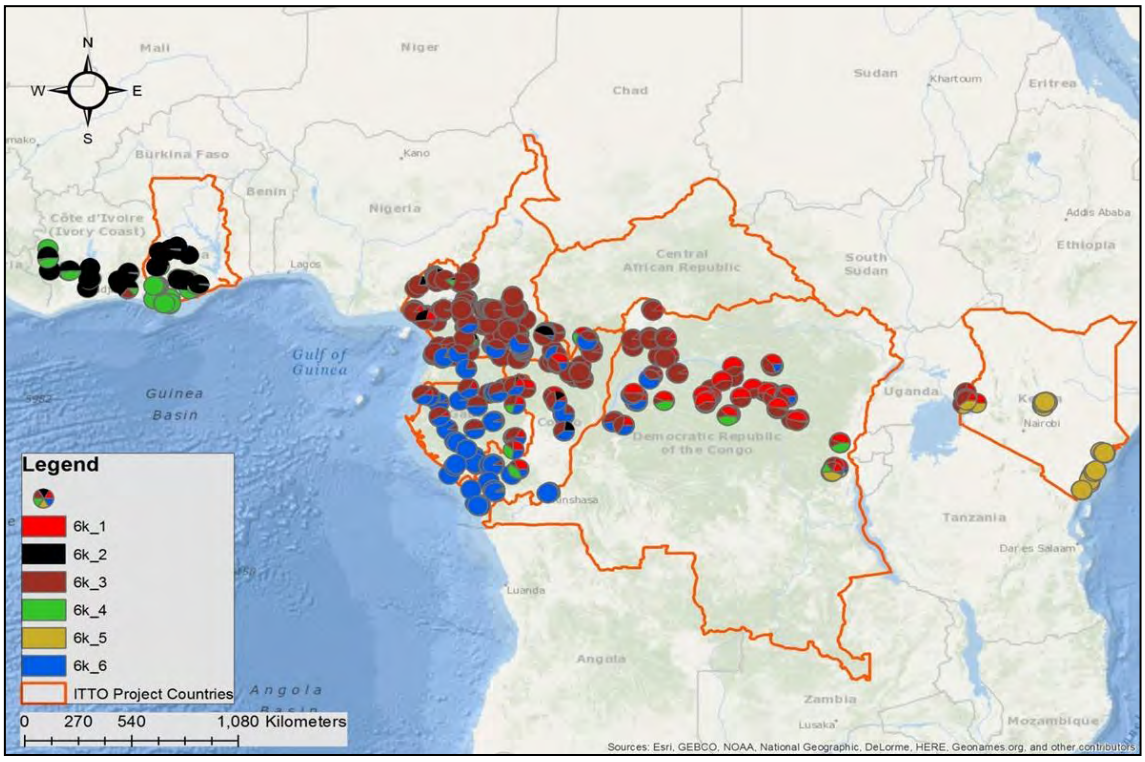


Figure 2*.Genetic reference data for Iroko (Milicia excelsa) in Africa. The colors visualize the genetic groups obtained with a Bayesian cluster analysis of 55 nuclear SNPS using the software STRUCTURE (Pritchard et al. 2000) for six putative genetic clusters.*

Ipe (*Handroanthus serratifolius*)

In frame of the Large-Scale project, we are developing in cooperation with partners in Brazil (Sao Paulo Forest Institute), Peru (IIAP), French Guiana (INRA) and Bolivia, a genetic reference database for the high value timber species Ipe (*Handroanthus sp*.). The actual reference data consist of 598 trees screened at a total of 122 SNPs (83 nSNPs, 13 cpSNPs, 25 mtSNPs). The combination of nuclear and plastid SNPs lead to a high success rate of self-assignments on the country level of 95% to 98% (figure 3).

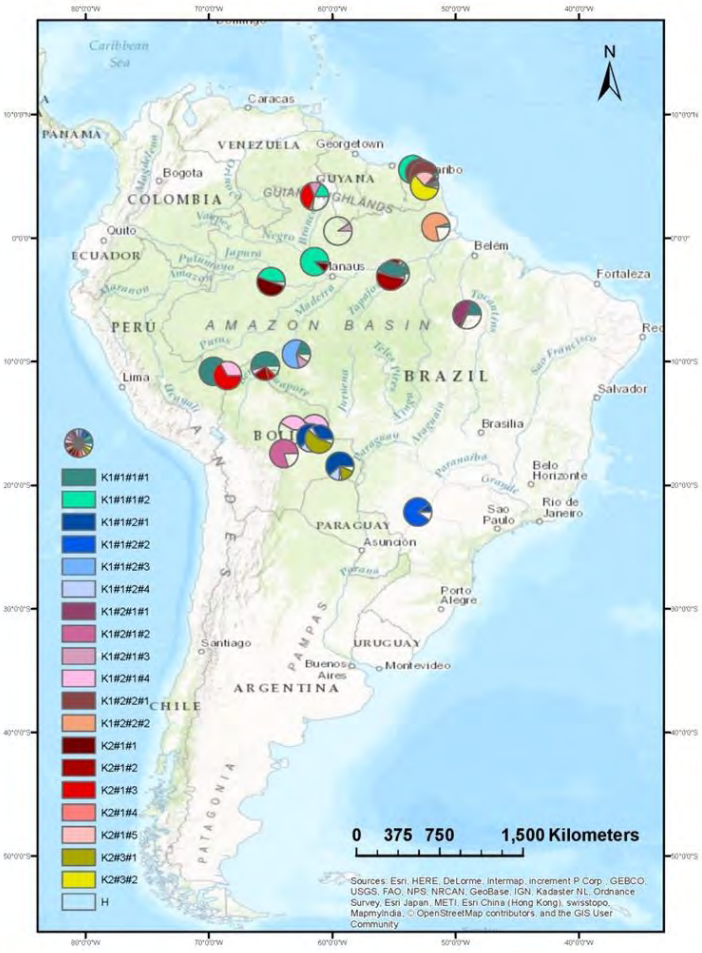


Figure 3*. Genetic reference data for Ipe (Handroanthus serratifolius) in South-America. The colors visualize the genetic groups and subgroups obtained with a hierarchical Bayesian cluster analysis of 121 SNPs using the software STRUCTURE (Pritchard et al. 2000).*

Practical applications

In March 2013, the Thünen Centre of Competence on the Origin of Timber was created. The Centre of Competence combines the expertise of the three Thünen Institutes of Wood Research, Forest Genetics, and International Forestry and Forest Economics, responsible for wood identification, proof of origin, certification and timber trade structures (http://www.thuenen.de/en/infrastructure/the-thuenen-centre-of-competence-on-the-origin-of-timber/). Since then, we received in the years 2013 to 2016 more than 1240 wood samples for genetic testing from competent authorities, non-governmental organizations (NGOs), timber traders and others (figure 4). In 10-20% of those cases, our tests identified false declarations on tree species or geographic origin.

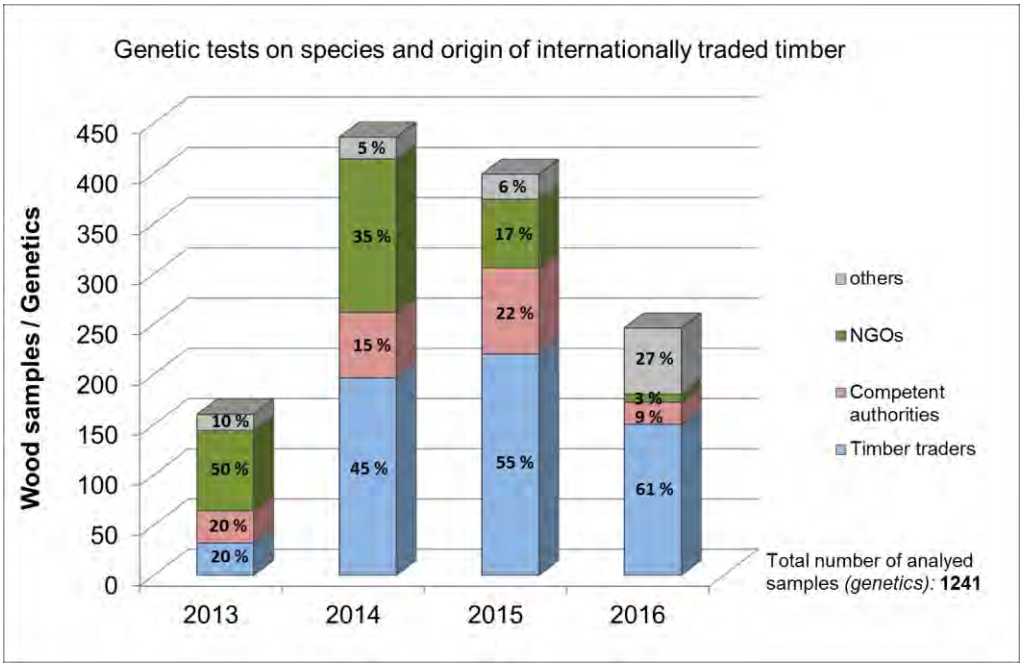


Figure 4. *Number of wood samples and classification of senders for genetic testing at the Thünen Centre of Competence on the Origin of Timber*

Conclusion and Outlook

In tree species, the genetic reference data often show cross boarder gene pools. Generally it is easy to genetically separate trees from different regions (groups of neighboring countries), also the genetic assignment to different populations is quite reliable. The medium geographic scale namely the country of origin is often a challenge. In addition, the species identification of reference samples caused for some species problems in the field or the biological separation of species is not strict. In these cases, there is in the genetic reference data often an overlapping of genetic signals (species differences, phylogeography, isolation by distance) that cause problems for the assignment of origin. Here, common data analysis approaches such as genetic assignments based on allele or haplotype frequencies have unsatisfying success rates. We recently developed an analysis method using a nearest neighbors approach which can improve the results (Degen *et al.* 2017). The best strategy is to group reference data according to the genetic clusters and to use them for testing of precise geographic claims. We have first promising results that show that genetic reference data of SNPs are better (have higher success rates of assignment) when they combine SNPs of different parts of the genome (nucleus, chloroplast, mitochondria). The genetic screening of reference material using multiplexed PCR-based techniques such as the MassARRAY approach is quite cost efficient and highly reliable. But the application of this approach for small numbers of wood samples is still too expensive. We need to look for alternative techniques for SNP genotyping on poor-quality DNA samples.

References

Besnard, G., F. Jühling, É. Chapuis, L. Zedane, É. Lhuillier, T. Mateille, and S. Bellafiore. 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode (*Meloidogyne graminicola*) using next generation sequencing. Comptes Rendus Biologies **337**:295-301.

Blanc-Jolivet, C., B. Kersten, N. Bourland, E. Guichoux, A. Delcamp, J.-L. Doucet, and B. Degen. 2017a. Development of nuclear SNP markers for the timber tracking of the African tree species Sapelli, Entandrophragma cylindricum. Conservation Genetics Resources:1-3.

Blanc-Jolivet, C., B. Kersten, K. Daïnou, O. Hardy, E. Guichoux, A. Delcamp, and B. Degen. 2017b. Development of nuclear SNP markers for genetic tracking of Iroko, Milicia excelsa and Milicia regia. Conservation Genetics Resources:1-3.

Dainou, K., C. Blanc-Jolivet, B. Degen, P. Kimani, D. Ndiade-Bourobou, A. S. L. Donkpegan, F. Tosso, E. Kaymak, N. Bourland, J. L. Doucet, and O. J. Hardy. 2016. Revealing hidden species diversity in closely related species using nuclear SNPs, SSRs and DNA sequences - a case study in the tree genus Milicia. Bmc Evolutionary Biology **16**.

Dainou, K., J. F. Flot, B. Degen, C. Blanc-Jolivet, J. L. Doucet, L. Lassois, and O. J. Hardy. 2017. DNA taxonomy in the timber genus Milicia: evidence of unidirectional introgression in the West African contact zone. Tree Genetics & Genomes **13**.

Degen, B., C. Blanc-Jolivet, K. Stierand, and E. Gillet. 2017. A nearest neighbour approach by genetic distance to the assignment of individual trees to geographic origin. Forensic Science International: Genetics **27**:132-141.

Degen, B., and H. Bouda. 2015. Verifying timber in Africa. ITTO Trop Forest Update **24**:8-10.

Gregorius, H.-R. 1987. The relationship between the concepts of genetic diversity and differentiation. Theoretical and Applied Genetics **74**:397-401.

Jardine, D. I., C. Blanc-Jolivet, R. R. M. Dixon, E. E. Dormontt, B. Dunker, J. Gerlach, B. Kersten, K. J. van Dijk, B. Degen, and A. J. Lowe. 2016. Development of SNP markers for Ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa. Conservation Genetics Resources **8**:129-139.

McKernan, K., C. Fujii, J. Ziauddin, J. Malek, and P. McEwan. 2002. A high throughput and accurate method for SNP genotyping using Sequenom MassARRAY (TM) system. American Journal of Human Genetics **71**:454-454.

Pakull, B., M. Mader, B. Kersten, M. R. M. Ekue, U. G. B. Dipelet, M. Paulini, Z. H. N. Bouda, and B. Degen. 2016. Development of nuclear, chloroplast and mitochondrial SNP markers for Khaya sp. Conservation Genetics Resources **8**:283-297.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics **155**:945-959.

Pujolar, J. M., M. Jacobsen, T. D. Als, J. Frydenberg, E. Magnussen, B. Jónsson, X. Jiang, L. Cheng, D. Bekkevold, and G. Maes. 2014. Assessing patterns of hybridization between North Atlantic eels using diagnostic single-nucleotide polymorphisms. Heredity **112**:627-637.

Chapter 4.2.

Opportunities for Improved Transparency in the Timber Trade through advanced DNA analysis

Andrew Lowe, Elly Dormontt, Darren Thomas
Centre for Conservation Science and Technology, University of Adelaide, Australia

Forests are important sources of timber, non-timber forest products, and other ecosystem services; tropical forests alone harbour more than half of the world's plant and wild animal species and store about 247 billion metric tons of carbon (Saatchi *et al.* 2011). Illegal logging is a major cause of forest degradation and subsequent loss (Burgess *et al.* 2012) estimated to account for between 15%–30% of the global trade in timber and worth US$30–$100 billion annually, including processing (Nellemann and INTERPOL 2012). In tropical regions, illegal logging rates are thought to be even higher, with 50%–90% of timber likely to be illegally sourced (Nellemann and INTERPOL 2012). The consequences of these illegal activities are realized economically, socially, and ecologically. Legitimate concession holders, governments, and local communities are denied vital revenue; armed conflict and corruption are promoted; and regional biodiversity assets and ecosystem services are degraded (Sikor and To 2011, Reboredo 2013).

Illegal logging for the international timber trade is predominantly a response to the external demand for wood products generated by consumer nations; therefore, efforts to curb the practice must address these demand drivers in addition to targeting illegal operations on the ground (Johnson and Laestadius 2011). In attempts to stem such international demand, legislation in Canada (1992), the United States (2008), the European Union (2010), and Australia (2012) now prohibits the importation of timber products harvested or traded in contravention of applicable foreign laws.

Nonstate market driven certification schemes have been developed in response to growing consumer demand for sustainable wood products and requirements to demonstrate compliance with timber regulations. Certification is obtained through initial assessment of compliance against a set of principles, criteria and indicators followed by periodic audits. Although it is difficult to fake compliance with standards of forest management and harvesting operations, the chain of custody of products along supply chains are vulnerable. Substitution or inclusion of prohibited timber, over harvesting, exclusion of sales from financial records and

mixing of certified and noncertified timber (Johnson and Laestadius 2011), present risks to the integrity of all certification schemes.

To tackle these concerns, in May 2014, the Member States of the United Nations recognized the need for "strengthening a targeted crime prevention and criminal justice response to combat illicit trafficking in forest products, including timber" (UNODC 2014). The resolution included the promotion of the development of tools and technologies that can be used to combat illicit trafficking of timber. Without the routine application of such verification tools, there can be little realistic expectation of demand-side initiatives significantly curbing the rates of illegal logging.

Current approaches to timber supply-chain verification

Reliance on paper-based methods alone leaves room for fraudulent activity. Documentation can be forged, or genuine documentation can be inappropriately associated with illegal timber. Efforts to implement more robust tracking systems using barcodes and electronic tagging go some way to ameliorating these risks (Seidel *et al.* 2012). However, the basic problem remains: Without a verification technique that derives from the timber itself rather than some externally affixed marker or associated paperwork, the system will always be vulnerable to the inclusion of illegal or otherwise non-compliant material. In order to genuinely verify that standards have been met, independent identification of the genus, species, geographic origin, specific individual, and, in some cases, the age of timber are required, based on characters inherent to the timber itself (Dormontt *et al.* 2015).

Scientific methods for timber supply-chain verification

Science can provide the means to identify timber, but it is not a trivial task. Timber does not have the most common diagnostic morphological features used for plant identification, such as flowers, fruits, and leaves. Therefore, the definitive scientific verification of timber has to rely solely on characteristics inherent in the wood itself. Various methods such as wood anatomical analysis (Wheeler and Baas 1998, Gasson *et al.* 2011), phytochemical analysis (Pastore *et al.* 2011, McClure *et al.* 2015), isotopic analysis (Kagawa and Leavitt 2010, Krüger *et al.* 2014), DNA barcoding (Lowe and Cross 2011, Jiao *et al.* 2015), and DNA profiling (Lowe *et al.* 2010, Jolivet and Degen 2012) are used to determine timber identity.

In this presentation, we have focussed on the use of genetic individualization for timber verification. Genetic individualization is the process of using the unique genetic profile of an individual to distinguish it from all others (excluding clones). The method is used extensively in human forensics to identify the origin of biological material. In timber identification, genetic individualization techniques can be used to verify whether shipments contain the same individuals at different points in the supply chain or whether there has been substitution or augmentation. Alternatively, the same techniques can be used to match timber evidence to the scene of illegal logging crimes. The technique is best suited to high-value timber, for which testing costs represent a lower fraction of the overall value of the timber and volumes and species diversity are typically low.

Genetic individualization to verify compliance in certified supply chains

In 2009, the International Tropical Timber Organization supported a project to evaluate the effectiveness of DNA verification of the chain of custody in CertiSource certified supply chains of Merbau timber (*Intsia* spp.) in Indonesia (Lowe *et al.* 2010, Seidel *et al.* 2012). Specimens were taken from logs at point of harvest in Papua and again on arrival at sawmills in Java. Genetic individualization was undertaken on a sample of matched specimens. The study revealed a DNA amplification success rate of between 59.2% (forest) and 41.9% (sawmill) and concluded that ongoing implementation of the system could be achieved at an affordable cost to industry. The application of scientific verification in this example can be used to demonstrate well-managed supply chains, and where mismatches are discovered, it can highlight weaknesses that can be further investigated by auditors.

Genetic individualization to identify illegal logging in US National Forest

In 2012, the US Forest Service uncovered sites of illegal logging of Bigleaf Maple (*Acer macrophyllum*) in the Gifford Pinchot National Forest. Timber off cuts from a nearby sawmill were seized as evidence. In a World Resources Institute–funded project, DNA markers (Jardine *et al.* 2015) and a subsequent DNA database were developed for the species that would provide individualization results suitable for admission to the US court system in support of a Lacey Act conviction (see table 2 for more information on the Lacey Act). The resulting database was used to test the evidence and revealed a highly significant match. All four defendants pleaded guilty in 2015–2016 (Dormontt et al, In Review). Research continues into reducing costs (see table 2 for cost details of the various methods) to enable the use of DNA verification in Bigleaf Maple supply chains, as well as for law-enforcement purposes.

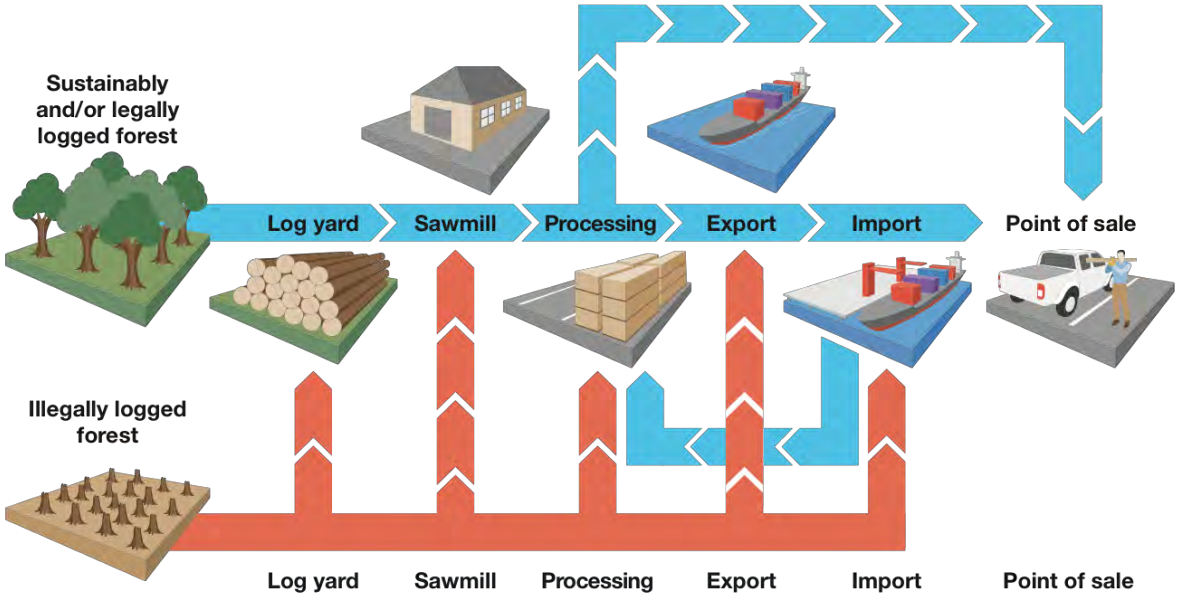
Scientific verification opportunities within the timber supply chain

The modern timber trade is characterized by complex global networks spanning multiple locations within producer nations and multiple consumer countries, making the challenge of monitoring and policing especially difficult. There are, however, discrete points along supply chains that present opportunities for routine scientific verification (figure 1). In most forests, standard management practices produce detailed inventories of standing trees. The collection of reference material to act as a benchmark for subsequent independent, scientific, supply-chain verification could be incorporated into the inventory process.

Requirements for implementation

The implementation of a global system of scientific timber supply-chain verification requires an integrated approach from policymakers, certification bodies, law-enforcement agencies, and industry. A concerted effort from the scientific community is also required to advance the development and forensic validation of identification technologies, to expand the scope of existing capabilities (more species, more geographic areas), and to continue to innovate in order to drive down costs. Certification systems have so far provided the only means through which consumers can make informed choices about wood product origins. However, the success to date of such schemes seems to present an unfortunate irony: The greater the consumer demand for certified products and the higher the prices consumers are

often willing to pay (Aguilar and Vlosky 2007), the greater the incentive for unscrupulous actors in the supply chain to defraud the system and reap the financial benefits of appearing to sell genuine certified products. Independent scientific verification embedded within existing certification schemes would provide the tools for certification bodies to police their supply chains, identify and exclude fraudulent products, and protect the integrity of their brand. Certification in other primary industries, such as fishing, has already begun to make such changes (MSC 2015), but beyond the pilot project of DNA verification of CertiSource products, timber certification schemes have so far steered clear of embedding scientific verification into their operating procedures.



	Log yard	Sawmill	Processing	Export	Import	Point of sale
Genus/Species ID	✓	✓	✓	✓	✓	✓
Source region ID			✓	✓	✓	✓
Individual tracking	✓	✓	✓			
Age determination				✓	✓	✓

Figure 1. *A schematic representation of the timber supply chain. Sustainably and/or legally harvested timber originates from appropriately managed logging concessions and is moved along the supply chain to log yards, saw mills, and processing plants. Products are then moved from processing to the point of sale or are exported for processing and reimported (often through multiple countries) before reaching the final point of sale. At each stage, illegally sourced timber products can enter the supply chain. A range of scientific technologies (visual, chemical, and genetic) exist that can be used to verify the legality of timber products at each stage of the supply chain.*

The routine use of timber-identification technologies by law-enforcement personnel policing trade routes would dramatically increase the rates of detection and prosecution of illegal logging crimes. However, implementation presents significant challenges: Distinguishing between legal and illegal timber is extremely difficult and requires access to experts and/or specialized tools. Law-enforcement agencies need to develop relationships with appropriate experts, raise awareness of the importance and availability of such resources, and train staff to select and acquire samples for testing. Given that timber is only a small part of their remit, the

resources to provide such support are likely beyond the reach of many law-enforcement agencies. Coordinated international efforts to address these needs present a potential solution. The International Consortium on Combating Wildlife Crime, a collaborative effort involving five intergovernmental organizations —the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), INTERPOL, the United Nations Office on Drugs and Crime (UNODC), the World Bank, and the World Customs Organization (WCO)— has convened an expert group, hosted by UNODC, bringing together customs and law-enforcement personnel, scientists, and legal professionals working on timber crime–related issues (UNODC 2015). The resulting guide, to be published this year, will detail how to acquire robust timber-identification outcomes.

Any implementation of routine timber-identification methods urgently requires increased investment and needs to direct effort towards the development and validation of scientific tests. Currently, the scientific basis of identification methods has been established, but the capacity for affordable routine testing in a wide range of taxa is generally lacking (Dormontt *et al.* 2015). A major impediment to the development of such tests is the paucity of taxonomically robust reference material from which identification methods and data can be derived. The current trend for reduced investment in collection-based science (Funk 2014) further impedes efforts to increase the pool of available timber-identification tests.

Outlook

Illegal logging is a complex global issue associated with a range of economic, social, and environmental drivers. The international scale of the problem demands an international response. Cooperation between timber producing, processing, and consuming nations is required and coordinated investment (both public and private) in scientific infrastructure. The technologies exist to encourage and enforce legal compliance, as well as improve sustainability, transparency, and consumer choice in the timber trade. Much work is still required, however, to expand the applicability of the available scientific verification methods and provide the policy, certification, and enforcement frameworks needed for effective routine implementation.

Acknowledgements

This article is rewritten from the text of the publication: Opportunities for Improved Transparency in the Timber Trade through Scientific Verification. AJ Lowe, EE Dormontt, MJ Bowie, B Degen, S Gardner, D Thomas, C Clarke, A Rimbawanto, A Wiedenhoeft, Y Yin, N Sasaki (2016) Transparency in the Timber Trade through Scientific Verification. BioScience 66: 990-998. This work was funded through ITTO grant no. TFL-PD 037/13 and ACIAR grants nos. FST/2014/028 and FST-2015-007 awarded to AJL. The authors gratefully acknowledge the US Department of Justice's Joseph Poux, the UK Border Agency's Guy Clarke, and Environment Canada's Jean-François Dubois for their constructive comments on table 1. The authors also thank Dr. Rob Ogden, Dr. Edgard Espinoza, Dr. Martin Breed, and the three anonymous reviewers for their constructive feedback, which greatly improved the quality of the manuscript.

References

Aguilar FX Vlosky RP 2007 Consumer willingness to pay price premiums for environmentally certified wood products in the US Forest Policy and Economics 9 1100 1112

Burgess R Hansen M Olken BA Potapov P Sieber S 2012 The political economy of deforestation in the tropics Quarterly Journal of Economics 127 1707 1754

Dormontt EE *et al.* 2015 Forensic timber identification: It's time to integrate disciplines to combat illegal logging Biological Conservation 191 790 798

Dormontt EE, Jardine D, Lowe AJ. Genetic profiling of timber provides tools for the prosecution of illegal logging crimes. Nature Sustainability. In review

Gasson P. 2011 How precise can wood identification be? Wood anatomy's role in support of the legal timber trade, especially CITES IAWA Journal 32 137 154

Jardine DI Dormontt EE van Dijk KJ Dixon RRM Dunker B Lowe AJ 2015 A set of 204 SNP and INDEL markers for Bigleaf maple (*Acer macrophyllum* Pursch) Conservation Genetics Resources 7 797 801

Jiao L Liu X Jiang X Yin Y 2015 Extraction and amplification of DNA from aged and archaeological *Populus euphratica* wood for species identification Holzforschung 69 925 931

Johnson A Laestadius L 2011 New laws, new needs: The role of wood science in global policy efforts to reduce illegal logging and associated trade IAWA Journal 32 125 136

Jolivet C Degen B 2012 Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon Forensic Science International Genetics 6 487 493

Kagawa A Leavitt SW 2010 Stable carbon isotopes of tree rings as a tool to pinpoint the geographic origin of timber Journal of Wood Science 56 175 183

Krüger I Muhr J Hartl-Meier C Schulz C Borken W 2014 Age determination of coarse woody debris with radiocarbon analysis and dendrochronological cross-dating European Journal of Forest Research 133 931 939

Lowe A Cross HB 2011 The application of DNA methods to timber tracking and origin verification IAWA Journal 32 251 262

Lowe A Wong K Tiong Y Iyerh S Chew F 2010 A DNA method to verify the integrity of timber supply chains: Confirming the legal sourcing of merbau timber from logging concession to sawmill Silvae Genetica 59 263 268

McClure PJ Chavarria GD Espinoza E 2015 Metabolic chemotypes of CITES protected Dalbergia timbers from Africa, Madagascar, and Asia Rapid Communications in Mass Spectrometry 29 783 788

[MSC] Marine Stewardship Council DNA Testing Assurance MSC. (19 August 2016 www.msc.org/about-us/ocean-to-plate-traceability/dna-testing-assurance)

Nellemann C & INTERPOL Environmental Crime Programme 2012 Green Carbon, Black Trade: Illegal Logging, Tax Fraud, and Laundering in the World's Tropical Forests: A Rapid Response Assessment United Nations Environment Programme, GRID-Arendal

Pastore TCM Braga JWB Coradin VTR Magalhães WLE Okino EYA Camargos JAA de Muñiz GIB Bressan OA Davrieux F 2011 Near infrared spectroscopy (NIRS) as a potential tool for monitoring trade of similar woods: Discrimination of true mahogany, cedar, andiroba, and curupixá Holzforschung 65 73 80

Reboredo F. 2013 Socio-economic, environmental, and governance impacts of illegal logging Environment Systems and Decisions 33 295 304

Saatchi SS *et al.* 2011 Benchmark map of forest carbon stocks in tropical regions across three continents Proceedings of the National Academy of Sciences 108 9899 9904

Seidel F Fripp E Adams A Denty I 2012 Tracking Sustainability: Review of Electronic and Semi-Electronic Timber Tracking Technologies International Tropical Timber Organization Technical Series no. 40

Sikor T To PX 2011 Illegal logging in Vietnam: Lam Tac (Forest Hijackers) in practice and talk Society and Natural Resources 24 688 701

[UNODC] United Nations Office on Drugs and Crime 2014 Resolution 23/1: Strengthening a Targeted Crime Prevention and Criminal Justice Response to Combat Illicit Trafficking in Forest Products, including Timber UNODC (19 August 2016; www.unodc.org/documents/commissions/CCPCJ/Crime_Resolutions/2010-2019/2014/Resolution_23_1)

[UNODC] United Nations Office on Drugs and Crime 2015 Outcome of the Expert Group Meeting on Timber Analysis (10–12 December 2014) Paper presented at the Commission on Crime Prevention and Criminal Justice Twenty-Fourth Session: World Crime Trends and Emerging Issues and Responses in the Field of Crime Prevention and Criminal Justice; 18–22 May 2015 Vienna, Austria (19 August 2016; www.unodc.org/documents/commissions/CCPCJ/CCPCJ_Sessions/CCPCJ_24/ECN152015_CRP4_e_V1503347.pdf)

Wheeler EA Baas P 1998 Wood identification: A review IAWA Journal 19 241 264

Chapter 4.3.

Challenges to implementation of high-throughput genotyping technologies for DNA forensics in the timber market

Stephen Cavers
Centre for Ecology & Hydrology, CEH Edinburgh, United Kingdom

Much has been made recently of the power of DNA-based approaches to deliver a system that can monitor timber movements and help to reduce the volume of illegally logged timber (Degen *et al.* 2013; Dormontt *et al.* 2015; Ng *et al.* 2016; Tnah *et al.* 2010; Jolivet and Degen 2012; Deguilloux *et al.* 2002). In particular the power of genomic (as opposed to marker-based genetic) technologies is being hailed as the way forward (Ogden 2011), with such methods able to deliver the greatly increased resolution needed to track timber movements in sufficient detail to permit enforcement. Allied to legal developments in major consumer markets and new frameworks to shape agreements between producers and consumer countries, it seems the time is ripe for major steps forward in control of illegal logging. However, whilst it is true that genomic tools have great potential to vastly improve detection of the geographic patterns of genetic structure that would underpin such a system, numerous challenges lie ahead. In particular, there is great urgency as forest - and particularly tree species of major market interest - continues to be lost at a significant rate. So these technological developments are taking place against a time limit, beyond which the resource we aim to protect will already be lost. In addition, accessibility to the technology and the gap between an operational system and reality on the ground remain major hurdles. In this paper, I focus on some of the challenges to implementation of an operational system using DNA forensics to monitor timber movements, focussing on the tropics as this is where the challenges are currently greatest (Anon n.d.). Although there is clear potential, several years since research efforts began to be trained on this issue, a working system has yet to emerge and it is valid to ask why this might be, as a means of drawing attention to the bottlenecks to achieving a successful system and to call for major investment to overcome them.

In the tropics, the rate of forest loss and speed of movement of the timber exploitation market should focus minds - headline examples make the case. Bigleaf Mahogany (Swietenia macrophylla), long a focus of conservation efforts underpinned by research (Degen *et al.* 2013; Grogan *et al.* 2014; Novick *et al.* 2003) and including CITES listed status (Snook 1996), is now effectively commercially extinct in many parts of its indigenous range. Despite efforts to protect it in both consumer and producer markets, exploitation repeatedly migrated to where the resource was least guarded and supply was maintained. Similar issues threaten other famous, internationally-traded species such as Brazilian rosewood (*Dalbergia nigra*), ipê (*Tabebuia spp.*), African mahogany (*Khaya spp.*), merbau (*Intsia palembanica),* as well as a long list of lesser known species.

An operational DNA forensic system needs to underpin confidence that owners can make use of resource within legally agreed framework. In essence this boils down to a couple of key things, reflecting the existing legal frameworks for sustainable exploitation of forest. Firstly, the systems needs to support assurance of trade in the declared species (is this the species I wanted to buy?); secondly it needs to support assurance that a wood product comes from the declared source (is this wood from where it says it is from?). The objectives of DNA forensics are therefore to verify species identity and to trace individual genotypes to their point of origin. Together with the question of whether the necessary checks to answer these questions can be delivered, this focuses the discussion on some key 'gaps'.

Diversity and taxonomic gap. DNA based species identification requires a database of verified species and a set of 'barcode' sequences for each species, which can be used to identify them. Currently, there is a limited set of verified barcodes for tropical species, but this is only part of the challenge. The raw species diversity in tropical forests is enormous and of this, the proportion of species used in the timber market is high. Estimates of the total number of tree species in the tropics varies between 35000 and 50000 (Slik *et al.* 2015); in the Amazon it is estimated at around 16000 (ter Steege *et al.* 2013). Hundreds of these are of commercial interest for timber. Even in an ideal world, the scale of the challenge of implementing DNA forensics for even a small proportion of the world's tropical timber species would be vast. However the problem is further complicated by the fact that not all tree species are distinct and evolutionarily well-separated; many remain poorly characterised from a taxonomic point of view (Cavers *et al.* 2013) , many may be hard to ever characterise robustly, being at early stages of phylogenetic diversification (Richardson *et al.* 2001), or part of species groups within which hybridisation readily occurs. Recent work on patterns of species diversity may offer routes to simplification of the problem (ter Steege *et al.* 2013) - Amazonia was recently shown to have strongly biased diversity (hyperdominance), with a few species accounting for the vast majority of the forest. This characteristic may allow prioritisation of species for genetic characterisation, with successful implementation of a system for a few species having a disproportionately large impact on the forest as a whole. Nevertheless, the rate of description of new species remains high, with even apparently well-known species subject to major taxonomic revision (State of the World's Plants, 2017; (Anon n.d.)). The taxonomic gap in tropical forests therefore presents a major challenge to setting in place a robust legal system.

There are some clear steps that could be taken to address the gap. Firstly, funding initiatives to support taxonomic description of tropical trees are essential but these should also target finding ways to accelerate the description of species. For example, whilst traditional botanical description proceeds, sequence databases could perhaps be compiled 'blind', with unstructured plant tissue collection supported by downstream phylogenetic analysis in the

way that current sequence-based 'microbiome' studies are revolutionising description of microbial communities. Secondly, reference collections need to be scaled up dramatically supported by a credible network of facilities for storage and access. This could perhaps be accelerated by linking sample provision to authority to log; the process of sample collection could overlay the description of a forest concession with a simple field protocol to encourage uptake and provision of datasets to sample providers for their own use. Finally, legal approaches need to take account of the fact that many species will never be fully described to the extent that they have clear boundaries, many will have 'fuzzy edges'. Building the flexibility to deal with such uncertainty needs to be part of the system, maybe best achieved by retaining a focus on what the outcomes of the system are meant to be (i.e. prioritise proof that a trader is working with sustainably managed, reference-identified product, rather than requiring proof that they have the internationally-agreed 'species X').

Sourcing and the spatial structure gap. Identification to source is perhaps the most novel aspect of the introduction of DNA forensics to tropical forest management. As trees are static and establish spatial genetic structure at various scales that reflects many factors including historical migrations, changes in population size, and contemporary ecology the potential exists for phylogeographic maps to provide the underpinning for a system of sourcing to geographic origin. In an ideal world, tree populations would be clearly bounded and distinct at a scale that reflects existing legal boundaries (e.g. concession, political border). However, as biological organisms, neither of these conditions are often likely to be met. Firstly, spatial structure generated by historical events is often only evident at large geographic scales, reflecting the scale of the processes that have driven it (Cavers and Dick 2013). Secondly, tropical tree ecology is characterised by the need to persist in highly diverse ecosystem where the density of conspecifics is likely to be low (Ward *et al.* 2005; Loveless and Hamrick 1984; Hamrick *et al.* n.d.). To overcome this, tropical tree species are often equipped to be highly dispersive, genes travel long distances and as a consequence population genetic structure is often weak. This means genetic groups may not fit the model required to enable source identification. For example, genetic groups may not be geographic (Rymer *et al.* 2013), where dispersal has driven simultaneous migration of multiple genetic groups or where mating is structured within populations. Or the scale of genetic structure may be incompatible with human legal or administrative boundaries. In Peru, for example, around 7 million ha of megadiverse tropical forest is now designated as forest concessions, with the average size of a concession is between 5-10000 ha (Salo and Toivonen 2009) but the spatial scale of genetic resolution for many species occurring in these forest areas is at a continental scale. Finally, the distinction between populations may simply be unclear. Ongoing long distance gene dispersal may simply preclude clear population boundaries and, whilst genetic differences may exist, they may be gradients of difference making the distinction of trees incompatible with a boundary-based administrative system. An additional potential complication is that, for some species, genetic structure may have been altered by human influence over historical time in which case patterns may either have been disrupted or may reflect encouragement of favoured genotypes.

Perhaps the clearest way forward for source identification strategies to become effective to attempt revise the system to reflect the resource, rather than the other way round. In other words, a species should be characterised and the administrative units should reflect the genetic structure that is detectable. Such a strategy would dovetail with other recently developed international proposals for management and conservation of Forest Genetic Resources, in that characterisation permits management (Loo *et al.* 2014).

The implementation gap. The final major gap is that between the state of the art in technology, and what is possible in the places where control is most urgently needed. In many producer countries, characterisation of forest resources is limited (also true in major consumer countries although the latter - being predominantly temperate - arguably have a more achievable task). Also, the capacity for analysis - either in terms of characterising tree species, or in application of forensic testing - is limited. Both would need major, sustained investment to allow development of a timber tracking system on a sufficient scale to have a substantial impact on the problem of illegal extraction and trading. Furthermore, in many producer countries, illegal resource exploitation issues extend beyond timber trees to many other tree species used in other ways. This includes internationally traded species such as *Prunus africana* (exploited for bark, (Vinceti *et al.* 2013) and *Osyris lanceolata* (used for incense, (Ooko 2009) but also locally exploited species - for example *Acacia senegal,* which is internationally valuable for gum arabic production (Odee *et al.* 2012) but is threatened by local use as firewood. In many cases, the priorities of producer countries do not mirror those of the major consumer countries so the question of which species are to be targeted for generation of genetic resources is not trivial.

Given that the hurdle of selecting species to target can be overcome, it is clear that major investment and support is necessary, particularly in producer countries, to establish and - vitally - to maintain capacity to characterise resources and to initiate testing procedures. In this process, it is essential that the producer countries are heard, supported and that outcomes reflect their local priorities as well as those of the international community.

Conclusions

Recognising the huge potential benefits of a well-resourced, scientifically underpinned system for DNA-based tracking of tropical timber, it is clear that a number of major challenges lie between the theory and making a system operational. It should also be borne in mind that for some species, due to evolutionary circumstance, population genetic history, or contemporary ecology it may never be feasible to monitor them using DNA methods. However, where the biological reality permits, a number of ways forward suggest themselves, which together would constitute major progress:

- Given the urgency of the need to act for tropical trees, rapid and large scale investment is needed to deliver an effective system quickly enough to have an impact on maintaining the resource. This needs to be across multiple aspects of the problem, addressing the taxonomic shortfall as well as in developing the practical aspects of widespread implementation of a technologically sophisticated system.
- It would be valuable to seek to develop a system of referenced databases as part of an international collaboration between producer and consumer countries, potentially supported by a credible internationally-recognised institution.
- Seek methods to scale up the rate of taxonomic description and population of barcode databases, in conjunction with prioritisation of species for in depth genetic resource characterisation. The latter should aim to distinguish local and international priorities but make room for both.
- Finally, support and investment in facilities and capacity in producer countries is essential and urgently needed, in a framework that sustains capability over the longer term.

References

State of the World's Forests 2016 | FAO | Food and Agriculture Organization of the United Nations [Online]. Available at: http://www.fao.org/publications/sofo/en/

State of the World's Plants 2017 | Royal Botanic Gardens, Kew [Online]. Available at: https://stateoftheworldsplants.com/ [Accessed: 15 November 2017b].

Cavers, S. and Dick, C.W. 2013. Phylogeography of Neotropical trees. *Journal of biogeography* 40(4), pp. 615–617.

Cavers, S., Telford, A., Arenal Cruz, F., Pérez Castañeda, A.J., Valencia, R., Navarro, C., Buonamici, A., Lowe, A.J. and Vendramin, G.G. 2013. Cryptic species and phylogeographical structure in the tree *Cedrela odorata* L. throughout the Neotropics. *Journal of biogeography* 40(4), pp. 732–746.

Degen, B., Ward, S.E., Lemes, M.R., Navarro, C., Cavers, S. and Sebbenn, A.M. 2013. Verifying the geographic origin of mahogany (*Swietenia macrophylla* King) with DNA-fingerprints. *Forensic science international. Genetics* 7(1), pp. 55–62.

Deguilloux, M.-F., Pemonge, M.-H. and Petit, R.J. 2002. Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proceedings. Biological Sciences / the Royal Society* 269(1495), pp. 1039–1046.

Dormontt, E.E., Boner, M., Braun, B., Breulmann, G., Degen, B., Espinoza, E., Gardner, S., Guillery, P., Hermanson, J.C., Koch, G., Lee, S.L., Kanashiro, M., Rimbawanto, A., Thomas, D., Wiedenhoeft, A.C., Yin, Y., Zahnen, J. and Lowe, A.J. 2015. Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biological Conservation* 191, pp. 790–798.

Grogan, J., Landis, R.M., Free, C.M., Schulze, M.D., Lentini, M. and Ashton, M.S. 2014. Big-leaf mahogany *Swietenia macrophylla* population dynamics and implications for sustainable management. *The Journal of applied ecology* 51(3), pp. 664–674.

Hamrick, J.L., Murawski, D.A. and Nason, J.D. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio*. Available at: https://link.springer.com/article/10.1007%2FBF00052230?LI=true.

Jolivet, C. and Degen, B. 2012. Use of DNA fingerprints to control the origin of sapelli timber (*Entandrophragma cylindricum*) at the forest concession level in Cameroon. *Forensic science international. Genetics* 6(4), pp. 487–493.

Loo, J., Souvannavong, O. and Dawson, I.K. 2014. Seeing the trees as well as the forest: The importance of managing forest genetic resources. *Forest Ecology and Management* 333, pp. 1–8.

Loveless, M.D. and Hamrick, J.L. 1984. Ecological Determinants of Genetic Structure in Plant Populations. *Annual review of ecology and systematics* 15(1), pp. 65–95.

Ng, K.K.S., Lee, S.L., Tnah, L.H., Nurul-Farhanah, Z., Ng, C.H., Lee, C.T., Tani, N., Diway, B., Lai, P.S. and Khoo, E. 2016. Forensic timber identification: a case study of a CITES listed species, *Gonystylus bancanus* (Thymelaeaceae). *Forensic science international. Genetics* 23, pp. 197–209.

Novick, R.R., Dick, C.W., Lemes, M.R., Navarro, C., Caccone, A. and Bermingham, E. 2003. Genetic structure of Mesoamerican populations of Big-leaf mahogany (*Swietenia macrophylla*) inferred from microsatellite analysis. *Molecular Ecology* 12(11), pp. 2885–2893.

Odee, D.W., Telford, A., Wilson, J., Gaye, A. and Cavers, S. 2012. Plio-Pleistocene history and phylogeography of *Acacia senegal* in dry woodlands and savannahs of sub-Saharan tropical Africa: evidence of early colonisation and recent range expansion. *Heredity* 109(6), pp. 372–382.

Ogden, R. 2011. Unlocking the potential of genomic technologies for wildlife forensics. *Molecular ecology resources* 11 Suppl 1, pp. 109–116.

Ooko, E.A.O. 2009. Evaluation of anti-microbial activity of *Osyris lanceolata* (East African Sandalwood). *Recent advances in forestry research for environmental conservation, improved livelihood and economic development. Proceedings of the 4th KEFRI Scientific Conference, KEFRI Headquarters, Muguga, Kenya, 6 to 9 October 2008.*. Available at: https://www.cabdirect.org/cabdirect/abstract/20103038517.

Richardson, J.E., Pennington, R.T., Pennington, T.D. and Hollingsworth, P.M. 2001. Rapid diversification of a species-rich genus of neotropical rain forest trees. *Science* 293(5538), pp. 2242–2245.

Rymer, P.D., Dick, C.W., Vendramin, G.G., Buonamici, A. and Boshier, D. 2013. Recent phylogeographic structure in a widespread 'weedy' Neotropical tree species, *Cordia alliodora* (Boraginaceae). *Journal of biogeography* 40(4), pp. 693–706.

Salo, M. and Toivonen, T. 2009. Tropical timber rush in Peruvian Amazonia: spatial allocation of forest concessions in an uninventoried frontier. *Environmental Management* 44(4), pp. 609–623.

Slik, J.W.F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L.F., Ashton, P., Balvanera, P., Bastian, M.L., Bellingham, P.J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F.Q., Breuer-Ndoundou Hockemba, M. and Venticinque, E.M. 2015. An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences of the United States of America* 112(24), pp. 7472–7477.

Snook, L.K. 1996. Catastrophic disturbance, logging and the ecology of mahogany (*Swietenia macrophylla* King): grounds for listing a major tropical timber species in CITES. *Botanical Journal of the Linnean Society* 122(1), pp. 35–46.

ter Steege, H., Pitman, N.C.A., Sabatier, D., Baraloto, C., Salomão, R.P., Guevara, J.E., Phillips, O.L., Castilho, C.V., Magnusson, W.E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J.C., Feldpausch, T.R., Coronado, E.N.H., Killeen, T.J., Mostacedo, B., Vasquez, R., Assis, R.L., Terborgh, J. and Silman, M.R. 2013. Hyperdominance in the Amazonian tree flora. *Science* 342(6156), p. 1243092.

Tnah, L.H., Lee, S.L., Ng, K.K.S., Faridah, Q.-Z. and Faridah-Hanum, I. 2010. Forensic DNA profiling of tropical timber species in Peninsular Malaysia. *Forest Ecology and Management* 259(8), pp. 1436–1446.

Vinceti, B., Loo, J., Gaisberger, H., van Zonneveld, M.J., Schueler, S., Konrad, H., Kadu, C.A.C. and Geburek, T. 2013. Conservation priorities for *Prunus africana* defined with the aid of spatial analysis of genetic data and climatic variables. *Plos One* 8(3), p. e59987.

Ward, M., Dick, C.W., Gribel, R. and Lowe, A.J. 2005. To self, or not to self... a review of outcrossing and pollen-mediated gene flow in neotropical trees. *Heredity* 95(4), pp. 246–254.

Chapter 4.4.

SNPs based timber tracking tools for African mahogany *Khaya* sp

Marius R. M. Ekué [a], ***Ulrich G. Bouka Dipelet*** [b,c], ***Birte Pakull*** [d], ***Blandine M. Y. Nacoulma*** [e], ***Emmanuel Opuni-Frimpong*** [f], ***Soulemane N. Yorou*** [g], ***Kudzo A. Guelly*** [h], ***Charles Doumenge*** [b], ***Judy Loo*** [i], *and* ***Bernd Degen*** [d]

[a] Bioversity International, Cameroon; [b] Forêts et Sociétés Unit, CIRAD, , France; [c] Laboratoire de Botanique et Ecologie, Université Marien Ngouabi, Congo; [d] Thünen Institute of Forest Genetics, Germany; [e] Laboratory of Plant Biology and Ecology, UFR-SVT, University of Ouagadougou, Burkina Faso / [f] University of Energy and Natural Resources, Ghana; [g] Faculty of Agronomy, University of Parakou, Benin; [h] Département de Botanique, Université de Lomé, Togo; [i] Bioversity International, Italy

The *Khaya* genus, also known as African mahogany, includes six species (*K. ivorensis, K. anthotheca, K. grandifoliola, K. senegalensis, Khaya nyasica* and *K. madagascariensis*) targeted by the illegal timber trade. Wood from the first three species is marketed together as "African mahogany" and is among the continent's most valuable sawnwood for export, with wide uses including shipbuilding. Policy instruments (e.g. EU Timber Regulation, Lacey Act) have been established to reduce illegal logging and associated trade at a global scale but practical control mechanisms to identify timber species and the geographic origin of wood and wood products are still lacking. The objectives were to develop a method for accurate species identification for African mahogany species, and to generate genetic reference data to identify the country of origin for the four main species (*K. ivorensis, K. anthotheca, K. grandifoliola,* and *K. senegalensis*) traded. 100 single nucleotide polymorphism (SNP) and one Indel markers were developed for *Khaya sp.* using a combination of restriction associated DNA sequencing and low coverage MiSeq genome sequencing. The markers were used to genotype 2,222 individuals of *Khaya* species collected in 18 African countries, using MassARRAY®iPLEX™ genotyping. Bayesian clustering produced four main genetic groups assigning all *K. ivorensis*, *K. senegalensis* and *K. grandifoliola* trees in three different clusters; and *K. anthotheca*, *K. madagascariensis* and *K. nyasica* in the fourth cluster. Discriminant analysis of principal component (DAPC) assigned individuals in 5 clusters with clear separation between *K. ivorensis*, *K. senegalensis*, *K. grandifoliola* and *K. anthotheca* trees. *K. madagascariensis* and *K. nyasica* were admixed. Genetic self-assignment tests with all 101 SNPs had success rates varying between 92% (*K. madagascariensis*) and 100% (*K. senegalensis*); except *K. nyasica* (62%). There was little evidence for hybridization among species and the vast majority (> 97%) of

individuals was assigned to the same species group as determined for them during sampling using morphological species classification. Genetic reference databases developed for dry zone mahogany (*K. senegalensis*) and broad-leaf mahogany (*K. grandifoliola*) from Ghana, Togo, Burkina Faso, Benin and Cameroon are presented. With *K. senegalensis*, the overall genepool differentiation among the five countries was delta δ = 0.10 and the fixation index F_{ST} = 0.18. The cluster analysis using the genetic distance among countries showed 4 main groups: Togo and Benin are less differentiated, while Burkina Faso, Ghana and Cameroun formed each a separate cluster with a bootstrap value ranging from 78 to 100%. The percentage of self-assignment of each individual back to its country of origin varies between 6 and 84% with the Bayesian approach (mean = 61%), and from 27 to 100% (mean=61%) with the nearest neighbor approach. With *K. grandifoliola*, the overall genepool differentiation among the six countries was delta δ = 0.17 and the fixation index F_{ST} = 0.22. The cluster analysis using the genetic distance among countries showed 2 sub-clusters with 100% bootstrap values: Togo and Benin are together in one cluster and Cameroon and Ghana in the second cluster. The percentage of self-assignment of each individual back to its country of origin varies between 29 and 100% (mean= 70%) with the Bayesian approach, and from 25 to 100% (mean 86%) with the nearest neighbor approach. Results of the blind tests showed clearly that the databases generated are robust and ready to be used to verify the origin of woods and wood products of *K. senegalensis* and *K. grandifoliola*; and for an accurate species identification of all African mahogany species.

Conclusions of PART 4

Application of DNA technologies to prevent timber illegal logging and trading

Based on learnt-lessons, it is critical to design different genotyping methods/coverage depending on the objectives (species ID, origin, individual fingerprinting) and the target species.

- - - - - - -

In addition, according to the final goal different genotyping strategies and use of standards should be applied, i.e. "quick" test to confirm/reject declarations (self-control of producers, timber traders, routine activities of law enforcement bodies and officials, etc) and forensic analysis.

- - - - - - -

Reference samples and standardized markers should be used to check the origin of the timber material. For this purpose, international cooperation in the frame of collaborative projects is required. This core funding should be reinforced by national funding support to fulfill the requirements of each country.

- - - - - - -

Complementary taskforces will be developed to discuss, propose and monitor the aforementioned activities.

FINAL OUTCOMES

Summary and proposed roadmap

International multidisciplinary collaboration is indeed required to address timber tracking, a complex task of global magnitude, which is threatening world-wide forests. Funding of future cooperation at scientific and technical level is strongly recommended to fulfill law enforcement needs by addressing the following requirements:

- *Develop an updated list of prioritized tree species based on GTTN's previous list. Identify reference samples from collections, enhancing collaboration with taxonomists, for sharing and transfer of reference material. Collect information on prioritized species from existing databases (with data pertinent to biology, geography, genome sequence or other molecular data availability, etc).*

- *Select different levels of testing according to the final aim of the analysis (barcoding to fingerprinting), considering the development of molecular markers for given geographic discriminations will involve high costs in research.*

- *Develop cost efficient new technologies for multispecies analysis and field-testing as required for law enforcement. This would also involve the assessment of the feasibility of applying 'old' data (e.g. microsatellites, AFLP markers) as often the newer technologies might not be the most cost-effective ones.*

- *Build a database, as a much-needed critical tool. New structures were discussed to integrate/harmonized existing resources (i.e. metadata-GTTN) with a centralized concept to capture data in a single-enriched database from different partners.*

Several task forces will be organized in collaboration with GTTN2 to accomplish the aforementioned requirements. To that end, funding instruments to build international collaborative actions are needed to integrate fragmented activities developed at world level. Additionally, national funding will complement these core activities to target national specific interests.