The KSA Native Genome Project

Nahed Mohammed [1], Ahmed Alghamdi [2], Ali Alabdali [2], Ikram Blilou[1], Heribert Hirt[1], Salim Al-Babili[1], Maged Saad[1], Khaled Alabdulkader [2], Rod Wing [1]

1 Center for Desert Agriculture, King Abdullah University of Science and Technology
2 National Center for Vegetation Cover Development & Combating Desertification, Ministry of Environment, Water and Agriculture

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

INTRODUCTION

Environmental protection *via* the conservation of biological diversity is a central theme of the Kingdom of Saudi Arabia's (KSA) Vision 2030. Biologically characterizing ecosystems that are relevant to the mission of MEWA's National Center for the Development of Vegetation Cover & Combating Desertification (NCDVC&CD) would provide essential data as to how and why a particular ecosystem operates in a particular environment. This information can then, in turn, be used to engineer large-scale ecosystems that are sustainable, will positively influence vegetation cover, and will fight against desertification across the KSA.

NATIVE GENOME "PILOT" PROJECT OBJECTIVES

1   Study a set of 15 terrestrial ecosystems, centered around 5 plant species, each falling into three categories – i.e. Threatened Species, Rangeland Species, and Common Species

2   Provide high quality reference genomes for each of these species and to investigate their ecosystems with respect to metabolomes, root development/architecture, and microbiomes

3   Establish a robust pipeline/infrastructure to characterize 100s-1000s of different ecosystems across the Kingdom, both terrestrial and aquatic, as part of a broader initiative entitled the KSA Native Genome Project (KSA-NGP) under development

4   Produce a new cohort of Saudi scientists trained in the study of ecosystem biology and its application to environmental protection and conservation

5   Promote sustainable agricultural practices throughout the kingdom and achieve food security in addition to protecting the environment by preserving biological diversity

LIST OF TARGETED PLANT SPECIES

The list includes common and endangered plant species of cultural, religious, medicinal and pastoral significance

Threatened species	Common species	Rangeland species
• *Mimusops laurifolia*	• *Juniperus procera*	• *Calligonum comosum*
• *Breonadia salicina*	• *Ziziphus spina-christi*	• *Leptadenia pyrotechnica*
• *Celtis africana*	• *Haloxylon persicum*	• *Salsola villosa*
• *Ficus vasta*	• *Salvadora persica*	• *Rhanterium epapposum*
• *Tamarindus indica*	• *Moringa peregrina*	• *Haloxylon salicornicum*

BRIEF PLAN OF WORK

1   Vouchering
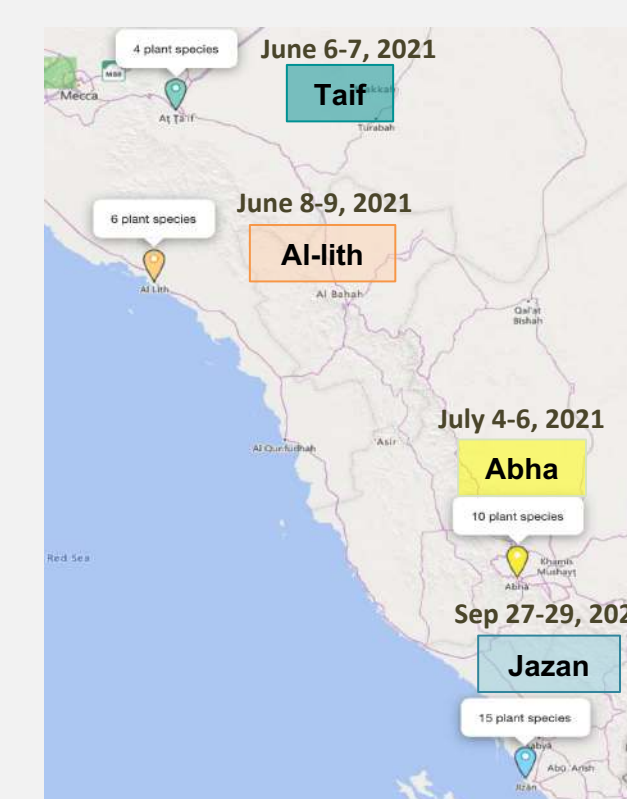
Sampling expeditions to identify living "Voucher" specimens

Plant tagging

Sample collection

Quality check and tissue harvest

2   Sequencing and data display and deposition

• **Plant genome sequencing**

A. Extract HMW DNA from leaf samples of each voucher specimen and perform QC analysis
B. Construct PacBio HiFi sequencing libraries, as well as libraries for Bionano optical mapping
C. Obtain baseline RNASeq data by sequencing RNA from young leaves, roots and mixed stage flowers of each voucher plant
D. Genome assembly and annotation following standard protocols established in the Wing Lab

• **Microbiome sequencing and analyses**

A. Obtain microbiome samples from roots and leaves of all voucher specimens for sequencing
B. Establish culture collections using standard protocols established in the Hirt lab
C. Genome assembly and annotation using Automatic Annotation of Microbial Genomes (AAMG)

3   Metabolomics data collection and analyses

A. Apply mass spectrometry (MS)-based plant metabolomics strategies to get a picture of primary and secondary metabolites in different tissues (i.e., leaves, roots, flowers or fruits)

4   Root development/architecture analyses

A. Establish the developmental profile for each of the proposed species over their entire life cycle
B. Analyze the root system's adaptive responses to optimize resource acquisition to thrive in different soil conditions
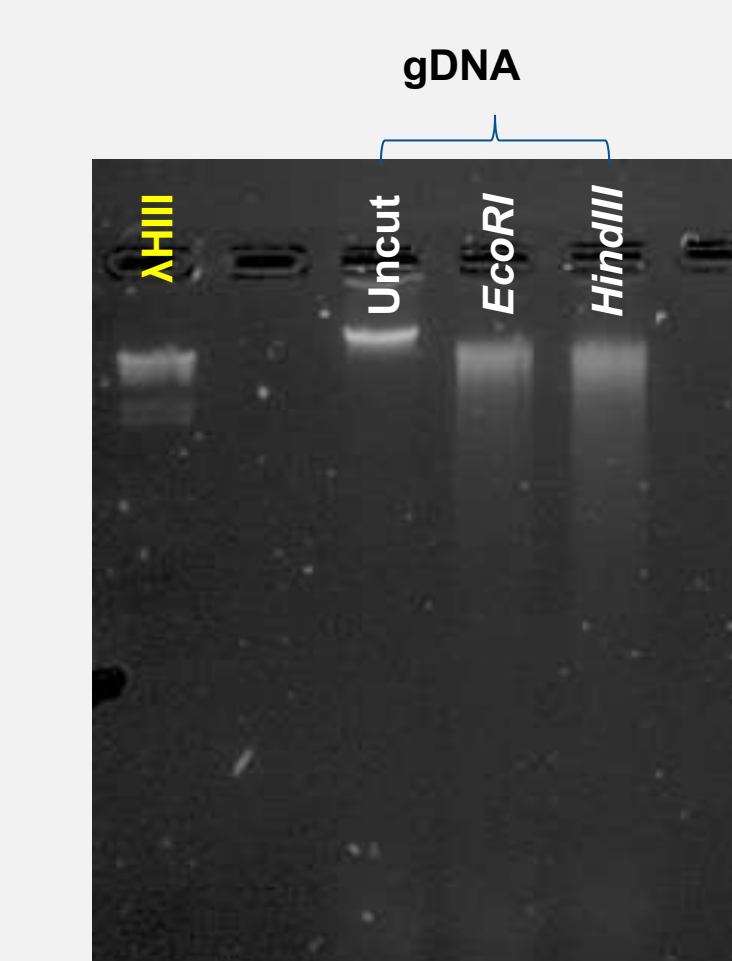
5   Data release

A. All sequence data, assemblies and annotations will be deposited in NCBI's Genbank for public access
B. Include all baseline metabolomics and root development/architecture data in KSA-NGP Ecosystem portal
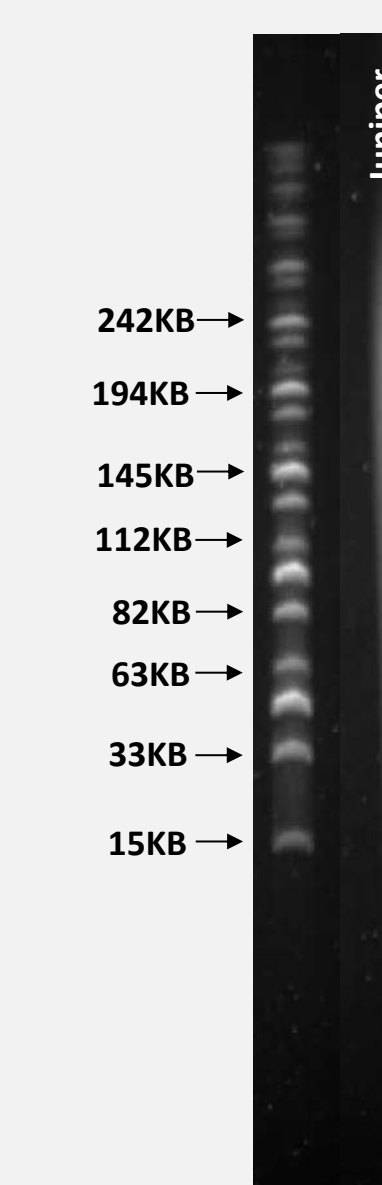
RESULTS

Example: *Juniperus procera*

• **DNA QC analysis**

A. Digestion QC   B. Size profile on CHEF   C. Quality and quantity check

Tissue (g)	4
Sample	*Juniperus procera*
260/280	1.88
260/230	1.97
Conc (ng/uL)	256
Vol (uL)	120
Total amount (ug)	31

• **Sequencing summary**

| Sample name	Sequencing platform	Sequence data	Coverage	Number of reads (M)	Average read length (Kb)	Genome Size
Juniperus procera	PacBio Sequel II- HiFi	~ 330 Gb	32X	>22	15.5	10.5 Gb

• **Genome assembly**

Total length of the assembly	No. of contigs	No. of contigs >1 Mb	Total length of contigs >1 Mb	Largest contig	N50 contig length	L50 contig count
11.2 Gb	8,800	2,227	9.9 Gb	46.6 Mb	5.6 Mb	554

PROGRESS AND FUTURE WORK

• A total of 44 species collected so far (including 15 targeted species and 29 additional species)
• A total of 9 targeted species was submitted for sequencing, 5 of which the sequencing data was released
• Arranging sampling expeditions to collect material from about 50-100 species
• Proceed with microbiome, metabolome and root development analyses for the collected species
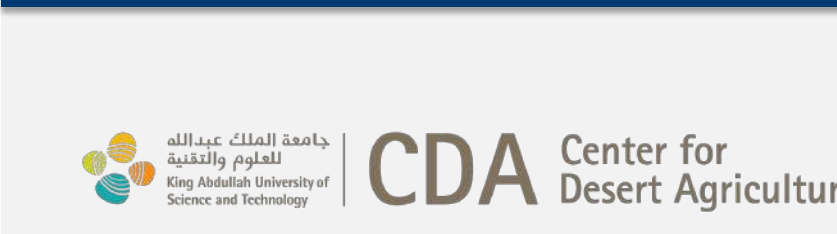
FURTHER INFORMATION

Nahed Mohammed

Nahed.mohammed@kaust.edu.sa