Clarification of *Elaeis guineensis* from NCBI databases

R Hayati¹, M Basyuni²*

¹Graduate School of Agroteknologi, Faculty of Agriculture, Universitas Sumatera Utara, Jl. Dr. A Sofyan No. 3 Medan, North Sumatera 20155, Indonesia
²Department of Forestry, Faculty of Forestry, Universitas Sumatera Utara, Jl. Tri Dharma Ujung No. 1 Medan, North Sumatera 20155, Indonesia

*Email: m.basyuni@usu.ac.id

**Abstract.** Oil palm is the essential oil-producing of the plant in the world. Two species genus in oil palm were *Elaeis oleifera* and *Elaeis guineensis* Jacq. The National Center of Biology Information (NCBI) contain databases of biology, related information about *E. Guineensis*, *E. Oleifera* and many other plant species. Identification of genes, DNA, proteins, active compounds and important taxonomies updated central server for researchers in the world. There are five numbers that have been a file in the source literature, included Bookshelf, Mesh, NLM Catalog, PubMed and PubMed Central literature. It found four books and reports and one Medical Subject Headings (MeSH), one NLM catalog reported, 346 PubMed and 1035 full-text journal articles about *E. guineensis* have been published. Furthermore, three assemblies, 43 BioProject, and 507 BioSample also available. Only exist one genome sequencing available in oil palm with 250,064 total DNA and RNA sequence *E.guineensis*. The NCBI collected 41,353 EST and 30,746 genes. There is also 51,353 groups of protein, and 30,746 sequences of protein have been reported.

1. Introduction

Elaeis consists of two species genus, specifically *Elaeis oleifera*, originally from America, and its twin species, namely *Elaeis guineensis* Jacq., from Africa, which is often referred to as African oil palm [1]. Oil palm is the most critical oil-producing of the plant in the world because of its high productivity; mostly the type is the African palm *Elaeis guineensis* Jacq [2]. New research on molecular and cellular studies of fruit abscess from tropical palm species *E. guineensis* and original functions and features of fruit absorption zones (AZ1) have been reported [3].

*E.guineensis* is very responsive to climate change, and environmental influences also intensify or reduce productivity [4]. Identification of genes, DNA, proteins, active compounds and important taxonomies in *E. guineensis* can be found at the National Center for Biotechnology Information (NCBI), an updated central server for researchers in the world. For example, information about species evolution and identification of important genes in *E. guineensis*, expression data for the R gene that identifies potential candidates for breeding disease resistance properties have been reported [5]. The study aims to report the application of the NCBI database (GQuery) to get more information about biotechnology on species *E. guineensis*.

2. Materials and method

The literature, health, genomes, genes, protein and chemical properties of *Elaeis guineensis* searched using NCBI databases (GQuery) (http://www.ncbi.nlm.nih.gov/). Accessed for database
available about oil palm plant. The calculated factors defined of the Bookshelf, NLM Catalog, PubMed, PubMed Central, OMIM, PubMed health, Assembly, BioProject, BioSample, Genome, Nucleotide, SRA, Taxonomy, PopSet, Identical Protein Groups, Protein, PubChem BioAssay, PubChem Compound, and PubChem Substance. Available searched online on November 7, 2018.

3. Results and Discussions

Table 1 shows the source literature available in the NCBI online related to *Elaeis guineensis*. Five numbers have been a file in the Bookshelf, Mesh, NLM Catalog, PubMed and PubMed Central literature. There are also four books and reports and one Medical Subject Headings (MeSH). On the other hand, one NLM catalog reported, and 346 PubMed and 1035 full-text journal articles about *E. guineensis* have been reported (Table 1). All of the databases publishing on internet technology are connected, can be used easily wherever [6].

| Literature     | Number | Description                        |
|----------------|--------|------------------------------------|
| Bookshelf      | 4      | Books and reports                  |
| MeSH           | 1      | Ontology used for Pubmed indexing  |
| NLM Catalog    | 1      | Books, journals and more in the NLM Collections |
| PubMed         | 346    | Scientific and medical abstracts/citations |
| PubMed Central | 1035   | Full-text journal articles         |

Information on health NCBI provides data retrieval and computational resource system for analyzing Medical Genetic (MedGen), an available one of data the latest development (Table 2).

| Health  | Number | Description                  |
|---------|--------|------------------------------|
| MedGen  | 1      | Medical genetics literature and links |

Table 3 shows the genomes of data is collected and maintained through several international collaborative researchers. Include three assemblies, 43 BioProject, and 507 BioSample. One genome sequencing project by organism *E. guineensis* completely, 1.8-gigabase (Gb) genome sequence with No Ref: GCF_000442705.1 [7]. Furthermore accumulated and maintained six genomes survey sequences, include 250.064 total DNA and RNA sequence *E. guineensis*.

One of which is Eukaryota (4-13 taxa) AFLP mentioned based Probe Set available reported and bioinformatics primers. The SRA resources described here 363, for example, Illumina HiSeq 2000 paired end of DNA sequencing. RNA-seq is useful for large-scale gene discovery, one of them related developmental mesocarp stages (miRNAs) have been reported [8]. The database functions all taxonomic nodes is one reported, from superkingdoms until subspecies, this search can be used to view the taxonomic positions of *E. guineensis*.

| Genomes     | Number | Description                      |
|-------------|--------|---------------------------------|
| Assembly    | 3      | genome assembly information      |
| BioProject  | 43     | biological projects providing data to NCBI |
| BioSample   | 507    | descriptions of biological source materials |
| Genome      | 1      | genome sequencing projects by the organism |
| GSS         | 6      | genome survey sequences         |
| Nucleotide  | 250,064| DNA and RNA sequences            |
| Probe       | 1,157  | sequence-based probes and primers |
| SRA         | 363    | high-throughput DNA and RNA sequence read the archive |
| Taxonomy    | 1      | taxonomic classification and nomenclature catalog |
Table 4 was shown to summarizes the number of genes available in NCBI. Expressed sequence tag (EST) in oil palm has been reported 41,353 and one of the cDNA array-based expression for the first time in several years ago [9]. Furthermore, the sources that lead to the locus of resistance was identified for Ganoderma disease in oil palm have been reported using random effects of the covariance matrix in the cox model and can be accessed online through this system [10]. The total 30,746 of the genes has been collected. The pattern of expression of certain genes related to *E. guineensis* can be accessed in all GEO datasets curated, 100 functional studies about genomic have been obtained. Whereas to examine the relationship of DNA sequences in certain oil palm populations can be found in the Popset menu.

| Genes source NCBI database for *Elaeis guineensis* | Number | Description |
|---|---|---|
| EST | 41,353 | expressed sequence tag sequences |
| Gene | 30,746 | collected information about gene loci |
| GEO DataSets | 100 | functional genomics studies |
| PopSet | 70 | sequence sets from phylogenetic and population studies |

Table 5 shows that many proteins belonging for oil palm plants available online in the database included 51,353 groups of protein and 30,746 sequences of protein. Recently, a green fluorescence protein gene for oil palm transgenic (GMO product) has been carried out but has not been a successful target [11]. On the other hand, the relationship between genes, protein and palm oil chemistry can be accessed through the BioSystem and was already available 385 of the total links, three of screening bioactivity and one deposited of the PubChem (Table 6). Some studies of the diversity of DNA profiles in the various commercial of Moderat Tahan Ganoderma (MTG) and three populations of pifisera type in *E. guineensis* used RAPD markers [12-13] have been reported.

| Proteins source NCBI database for *Elaeis guineensis* | Number | Description |
|---|---|---|
| Identical Protein Groups | 41,353 | protein sequences grouped by identity |
| Protein | 30,746 | protein sequences |

| Chemicals source NCBI database for *Elaeis guineensis* | Number | Description |
|---|---|---|
| BioSystems | 385 | molecular pathways with links to genes, proteins, and chemicals |
| PubChem BioAssay | 3 | bioactivity screening studies |
| PubChem Substance | 1 | deposited substance and chemical information |

The National Center of Biology Information (NCBI) contain databases of biology, related information about *E. Guineensis, E. Oleifera* and many other plant species. The web is managed by an institution domiciled in the United States. NCBI was a program on the dream of revealing to new knowledge, especially in the molecular field. NCBI provides several interesting facilities for molecular researchers by opening an information portal that is mostly free.

4. Conclusions
The online library of National Center Biotechnology and Information (NCBI) has been clear about *Elaeis guineensis*. All of the databases updated the information and can change at any time. The NCBI of tools supported that provide updates several researchers regarding *E. Guineensis*. 
Elaeis guineensis Jacq. - the palm (Elaeis, Elaeis, Elaeis, Elaeis, Elaeis) fruit abscission: A model for studying fruit

text continues...