Acta Universitatis Agriculturae Sueciae

Doctoral Thesis No. 2023:100
Faculty of Landscape Architecture, Horticulture and Crop Production Science

Nutritional profiling, SNP-based genomic tools development for population genetic analysis and genotype-by-environment interaction in noug (*Guizotia abyssinica* (L.f.) Cass.)

Sewalem Tsehay Wondim

Nutritional profiling, SNP-based genomic tools development for population genetic analysis and genotype-by-environment interaction in noug (*Guizotia abyssinica* (L.f.) Cass.)

Sewalem Tsehay Wondim
Faculty of Landscape Architecture, Horticulture and Crop Production Sciences
Department of Plant Breeding
Alnarp

SLU
SWEDISH UNIVERSITY
OF AGRICULTURAL
SCIENCES

DOCTORAL THESIS
Alnarp 2023

Acta Universitatis Agriculturae Sueciae
2023:100

Cover design/photo: The front cover image proceeds in a clockwise direction, beginning at the right corner with image of noug seeds, pre-flowering noug crop, flowering stage of noug crop, the by-product of cold-pressed noug seed, a partial view of the primary fatty acids in noug oil, noug seed oil packed by industry (https://www.indiamart.com/proddetail/niger-seed-oil-23315445973.html) and at the center lies packaged noug seed oil.
(Design/photo: Sewalem Tsehay Wondim)

ISSN 1652-6880
ISBN (print version) 978-91-8046-252-5
ISBN (electronic version) 978-91-8046-253-2
https://doi.org/10.54612/a.3sj19brqcf

Swedish University of Agricultural Sciences, Department of plant breeding, Alnarp, Sweden
Print: SLU Service/Repro, Alnarp 2023

Abstract

Noug (*Guizotia abyssinica*) is an economically important oilseed crop and has its primary center of origin and diversity in Ethiopia. It is mainly cultivated in Ethiopia and India. Noug is a self-incompatible dicotyledonous annual herbaceous semi-domesticated crop. The nutritional content of diverse germplasm of noug seed was evaluated for total lipid, fatty acids, protein and 12 macro and micro-mineral elements using standard bioanalytical techniques. Noug seed contains 33–49% lipid with rich linoleic acid, an essential polyunsaturated omega-6 fatty acid (72–79%) and consequently low in oleic acid (5–9%). The total protein content of noug seeds ranged from 25 to 28%. Significant differences in mineral element contents were revealed across experimental locations with potassium (K) being the highest 9254 ($\mu g\ g^{-1}$) in concentration. Meeting the demands of the world´s growing population in the presence of climate change demands intensive crop breeding. Breeding oilseed crops is part of the race to meet consumers' rising demands. However, before an intensive breeding program is initiated, we have to lay the foundation with pre-breeding. Genomic tools facilitate our deep insights into crop genetic diversity, population structure, genotyping, and marker-assisted selection, which are valuable for breeding and conservation programs. This research developed novel DNA markers for noug based on transcriptome sequencing. For this, 959 single nucleotide polymorphisms (SNPs) were identified from 628 transcripts of two noug genotypes. Thereafter, kompetitive allele-specific PCR (KASP) markers were developed based on these SNPs and utilized in genotyping noug accessions. The genotypic data were utilized for population genetic analyses. The SNP loci had an average of 0.24 polymorphic information content (PIC) and about 50% of the loci showed significant deviation from Hardy-Weinberg equilibrium. The markers revealed a high genetic variation within accessions. The developed genomic resources proved useful for population genetic diversity analysis. Breeding noug involves improving its oil content, quality, and productivity, as well as adaptability to diverse environments. This research also evaluated the magnitude of the effect of genotype-by-environment interaction and stability of noug nutritional value traits including oil content, linoleic, oleic, palmitic and stearic acids, and the results provided useful insights, which could facilitate the development of superior cultivars through breeding.

Keywords: Genetic diversity, Genotype-by-environment interaction, Genotyping, *Guizotia abyssinica*, KASP markers, Lipids, Minerals, Noug, Oilseeds, Population structure, Protein, SNPs

Näringsanalys, utveckling av SNP-baserade genomiska verktyg för populationsgenetisk analys och genotyp-miljö-interaktion hos noug (Guizotia abyssinica)

Sammanfattning

Noug (*Guizotia abyssinica*) är en ekonomiskt viktig oljeväxt som har sitt huvudsakliga ursprung och sin primära diversitet i Etiopien. Den odlas främst i Etiopien och Indien. Noug är en självinkompatibel tvåhjärtbladig årlig ört som är semi-domesticerad. Den näringsmässiga sammansättningen av olika accessioner av noug utvärderades för fetthalt, fettsyrasammansättning, proteinhalt samt 12 makro- och mikromineralämnen, med hjälp av bioanalytiska standardmetoder. Nougfrön innehåller 33–49% fett med hög andel linolsyra, en essentiell fleromättad omega-6-fettsyra (72–79%) och låg oljesyrahalt (5–9%). Det totala proteininnehållet i nougfrön varierade från 25 till 28 %. Signifikanta skillnader i mineralinnehållet från frön som odlats i olika försöksplatser observerades, där kalium (K) uppvisade högsta värdet med en koncentration på 9254 μg g^{-1}. Att möta behovet av världens växande befolkning i en tid av klimatförändringar kräver intensivt växtförädlingarbete. Förädling av oljeväxter är en viktig del för att möta konsumenternas ökande krav. Innan ett intensivt förädlingsprogram initieras måste dock grunden läggas med för-förädlingsinsatser (pre-breeding). Genomiska verktyg möjliggör djupare förståelse i grödans genetiska diversitet, populationsstruktur, genotypning och markörbaserad urval, vilket är värdefullt för förädlings- och bevarandeprogram. Inom ramarna för denna forskning utvecklades nya DNA-markörer för noug baserat på transkriptomdata. Genom detta identifierades 959 SNP från 628 transkript av två nouggenotyper. Baserade på dessa SNP:er utvecklades sedan kompetitiva allelspecifika PCR (KASP)-markörer och användes för genotypning av nougaccessioner. Genotypdata användes för populationsgenetiska analyser. De identifierade SNParna hade i genomsnitt ett värde på 0,24 för polymorfic information (PIC), och cirka 50% av SNParna visade signifikant avvikelse från Hardy-Weinbergjämvikt. Markörerna visade en hög genetisk variation inom accessioner. De utvecklade genomiska resurserna visade sig vara användbara för analys av populationsgenetisk diversitet. Förädling av noug innebär att förbättra dess oljeinnehåll, kvalitet och produktivitet samt anpassning till olika miljöer. Denna forskning utvärderade också omfattningen av effekten av genotyp-miljö-interaktion samt stabiliteten av näringsvärdet i noug, inklusive oljeinnehåll, linolsyra, oljesyra, palmitinsyra och stearinsyra, och resultaten resulterade i insikter som kan bidra till utvecklingen av förbättrade sorter genom växtfördlingsinsatser.

Nyckelord: Genetisk diversitet, Genotyp-miljö-interaktion, Genotypning, *Guizotia abyssinica*, KASP-markörer, Lipider, Mineraler, Noug, Oljeväxter, Populationsstruktur, Protein, SNP.

Preface

The work detailed in this PhD thesis summary represents collaborative developmental research involving the Swedish University of Agricultural Sciences, Addis Ababa University, and the Ethiopian Institute of Agricultural Research, conducted across both Ethiopia and Sweden. Within this thesis, two published articles and one manuscript are included, wherein the PhD student took the lead role. The main works that the PhD student encompassed designing and executing field experiments, doing laboratory experiments, overseeing greenhouse activities, conceptualizing methodologies, analysing data using specific software, drafting papers, and collaborating with co-authors in reviewing and editing paper outputs emanated from this research among others. This summary, more than a mere abstract, serves as a comprehensive synthesis encapsulating the core of the PhD thesis.

Dedication

The dedication of this work is to my spouse, Rekik Tizazu, and my children, Matias Sewalem, Natanim Sewalem, and Henok Sewalem.

I also dedicate this doctoral work to the memory of the esteemed associate professor Kifle Dagne (PhD), my former teacher and colleague at Addis Ababa University. He exemplified unwavering dedication as both an instructor and a researcher at Addis Ababa University from 1971 to 2020. The photograph below displays the late Kifle Dagne's image.

ትጥቄ ከሚላላ ጎጃም ያስታጠቀኝ ፣
በላይ የሞተበት ገመዱ ይነቀኝ።።
(ከሙዚቀኛ መሃሪ ደገፋው ሸለመኝ ከሚለዉ የሙዚቃ ግጥም የተዋሰደ)

Contents

List of publications

The foundation of this doctoral thesis lies in the content of the subsequent papers, distinguished by Roman numerals as listed below and referenced in the text when needed:

I. **Tsehay, S**., Ortiz, R., Geleta, M., Bekele, E., Tesfaye, K. & Johansson, E. (2021). Nutritional Profile of the Ethiopian Oilseed Crop Noug (*Guizotia abyssinica* Cass.): Opportunities for Its Improvement as a Source for Human Nutrition. Foods, 10(8), 1778. https://doi.org/10.3390/foods10081778.

II. **Tsehay, S**., Ortiz, R., Johansson, E., Bekele, E., Tesfaye, K., Hammenhag, C. & Geleta, M. (2020). New Transcriptome-Based SNP Markers for Noug (*Guizotia abyssinica*) and Their Conversion to KASP Markers for Population Genetics Analyses. Genes, 11(11), 1373. https://doi.org/10.3390/genes11111373.

III. **Tsehay, S**., Geleta, M., Bekele, E., Tesfaye, K. & Johansson, E., Tesfaye, M., Ortiz, R. Genotype-by-environment interaction and stability analysis of seed oil content and major fatty acids in noug (*Guizotia abyssinica* (L.f.) Cass.). (Manuscript).

The contributions of Sewalem Tsehay Wondim to the papers included in this thesis were as follows:

I. Conceptualization was conducted by M.G., E.B., R.O., and K.T.; methodology was developed by E.J. and M.G.; software was handled by S.T. and E.J.; validation involved S.T., E.J., and M.G.; formal analysis was conducted by S.T. and E.J.; the investigation was carried out by S.T. and M.G.; resources were managed by M.G.; data curation was performed by M.G.; the original draft preparation was written by S.T.; review and editing were completed by R.O., E.J., E.B., K.T., and M.G.; visualization tasks were managed by S.T. and E.J.; supervision was provided by R.O., E.J., M.G., E.B., and K.T.; project administration was overseen by R.O. and M.G.; and funding acquisition was coordinated by M.G., E.B., R.O., K.T., and E.J.

II. The conceptualization was led by M.G., R.O., and S.T.; methodology was crafted by M.G., R.O., and S.T.; data analysis involved S.T., M.G., and C.H.; the original draft preparation was written by S.T.; review and editing were carried out by M.G., R.O., E.J., E.B., K.T., C.H., and S.T.; funding acquisition was managed by M.G., E.B., K.T., E.J., and R.O.; and supervision was provided by R.O., M.G., E.J., E.B., and K.T.

III. S.T. was responsible for designing and executing field and laboratory experiments, conducting data analysis, and composing the manuscript, while R.O., M.G., E.J., E.B., K.T., and M.T. contributed to reviewing the manuscript.

List of tables

List of figures

Abbreviations, acronyms and symbols

AMOVA	Analysis of molecular variance
cpDNA	Chloroplast DNA
DNA	Deoxyribonucleic acid
EST-SSRs	Expressed sequence tag-derived simple sequence repeats
F	Fixation index
FAMEs	Fatty acid methyl esters
GC	Gas chromatography
GEI	Genotype-by-environment interaction
He	Expected heterozygosity
Ho	Observed heterozygosity
Hs	Nei's gene diversity
HWE	Hardy-Weinberg equilibrium
I	Shannon diversity index
ITS	Internal transcribed spacer
KASP	Competitive allele-specific PCR
MAF	Minor allele frequency
N	Nitrogen
Nm	Gene flow
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
PhD	Doctor of Philosophy
PIC	Polymorphism information content

PPL	Percent polytrophic loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
SLU	Swedish University of Agricultural Sciences
SNPs	Single nucleotide polymorphisms
TAG	Triacylglycerol

1. Introduction

The human population is expected to reach 10 billion in 2050 (Sagun, Yadav & Alonso 2023). As the world population continues to grow, the demands for renewable and sustainable food and energy have increased. This has triggered intensive research and development programs locally and globally. Specifically, the surge in edible vegetable oils demand has been a key motivating factor for breeding oilseed crops to improve oil yield and quality. The consumption of vegetable oil from seed oil crops has increased with the growing population, contributing to the improvement of quality of life (Sharma, Gupta & Mondal 2012; Samarth & Mahanwar 2015; Ahmad *et al.* 2021). Triacylglycerols (TAGs) are plant lipids composed of three fatty acids linked to the glycerol backbone and are mainly stored in oilseeds (Gunstone 2011) as shown in Figure 1.

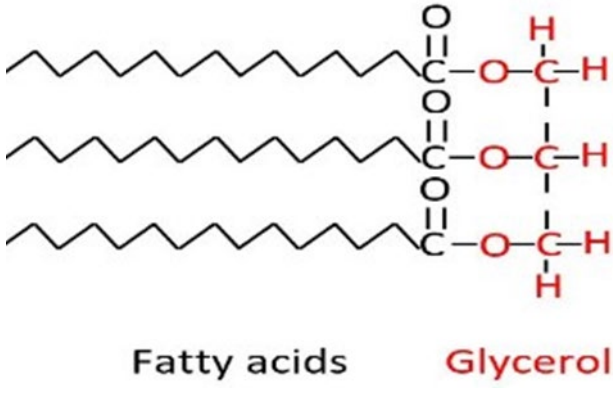


Figure 1. Chemical structure of a triacylglycerol (TAG).

Triacylglycerols are the main components of vegetable oil which are renewable resources for animal diets and several industrial applications (Durrett, Benning & Ohlrogge 2008; Msanne, Kim & Cahoon 2020; Subedi *et al.* 2020). Vegetable oils are the main reservoirs of calories and major sources of essential fatty acids (Kumar, Sharma & Upadhyaya 2016). The demand for these commodities has increased significantly in recent years, both in terms of food for humans and animals as well as for industrial applications (Frančáková *et al.* 2015; Kumar, Sharma & Upadhyaya 2016; Nieto & Lorenzo 2022). Vegetable oils can be obtained from seeds, fruits, germs (seed embryos), leaves, and so on in different amounts and compositions of their constituent fatty acids (Lee *et al.* 2017; Zhou *et al.* 2020). Vegetable oils are major sources of monounsaturated and polyunsaturated fatty acids as well as sources of fat-soluble vitamins and antioxidants (Savva & Kafatos 2016). Fatty acid composition of vegetable oil is a major determinant of the nutritional quality and health-promoting attributes of vegetable oils (Msanne, Kim & Cahoon 2020).

Oilseed crops have gained paramount importance to agricultural production over the past 40 years as useful commodities worldwide due to the growing demand for vegetable oils and protein means (Attia *et al.* 2021). Considering the increasing demand for high-quality edible oils, research on oilseed crops has been directed towards increasing oil content and quality (Li *et al.* 2022). Different types of research approaches including crossbreeding, molecular breeding, and targeted genetic modifications have been used to achieve these breeding goals (Haroon *et al.* 2020).

Oil quality is determined by the composition and proportion of fatty acids (Deme *et al.* 2021). Proper development of molecular and genomic tools for the characterization of genetic resource and the phenotyping of fatty acid composition and oil content of diverse genetic resources facilitate the improvement of targeted oil crops (Porokhovinova *et al.* 2022).

Several oilseed crops are grown in Ethiopia. Among them, noug (*Guizotia abyssinica* (L. f.) Cass.), linseed (*Linum usitatissimum* L.), Ethiopian mustard (*Brassica carinata A. Braun)*, sesame (*Sesamum indicum L.*), groundnut

(*Arachis hypogaea* L.), sunflower (*Helianthus annuus* L.), castor (*Ricinus communis L.*), soybean (*Glycine max (*L.*) Merr*.), and cotton (*Gossypium hirsutum* L.) are economically significant. Noug and Ethiopian mustard are indigenous to Ethiopia (Alemaw & Gurmu 2023).

Noug is a minor and research-neglected oilseed crop with significant economic potential in Ethiopia and India (Agegnehu 2010; Tadele 2018; Palchamy *et al.* 2022). In spite of its significant potential for food security, especially in Ethiopia, as a source of edible oil, protein, and minerals as well as animal feed, noug has not received adequate research and breeding efforts for improvement. Research in genomic resources plays a crucial role in the effective manipulation and utilization of genetic resources for the improvement of noug germplasm. Genomic resources provide valuable tools and information to achieve this goal (Genetics 2020; Nair & Pandey 2021; Mir *et al.* 2022). Genomic resources are used in genetic diversity analysis, genome sequencing, gene discovery, marker-assisted breeding, Identification of quantitative trait loci (QTL), transcriptomics, proteomics and functional genomics among others (Mohd Saad *et al.* 2021; Thudi *et al.* 2021; Varshney *et al.* 2021). By integrating these genomic resources, scientists can make informed decisions and apply advanced techniques to enhance the agronomic, nutritional, and economic value of noug. This approach contributes to crop improvement, ensuring its sustainability and resilience in the face of changing environmental conditions and economic demands. Such genetic resources are not sufficiently available for noug. Hence, one of the major aims of this thesis research is to contribute to the development of genetic resources for noug. This approach contributes to crop improvement, ensuring its sustainability and resilience in the face of changing environmental conditions and economic demands.

The main aims of this PhD thesis research were to access nutritional profiles, develop genomic tools for variability evaluation of noug genetic resources, and estimation of the effect of genotype by environment interaction in the oilseed crop noug in Ethiopia. The result of this research has practical implications for breeding and genetic improvement through the utilization of noug genetic resources in Ethiopia.

2. Background

2.1 Taxonomy and phylogenetics of the genus *Guizotia*

The family Asteraceae is the largest angiosperm family comprising circa 25,000 species in 1500-1700 genera, with worldwide distribution except Antarctica (Nie *et al.* 2016; Giberti 2018; Zhang *et al.* 2021). *Guizotia* is a small genus in the tribe Heliantheae of the family Asteraceae. During her revision of the taxonomy of the genus *Guizotia* in 1974, Baagøe placed the genus within the subtribe Coreopsidinae (Baagøe, 1974). However, a phylogenetic analysis of various subtribes of the tribe Heliantheae strongly supports its placement within the subtribe Milleriinae (Geleta et al. 2010). Baagøe's revision of the genus in 1974 resulted in six species, one of which comprising two subspecies. These are *G. abyssinica* (L. f.) Cass., *G. arborescens* I. Friis, *G. jacksonii* (S. Moore) J. Baagøe, *G. scabra* (Vis.) Chiov. ssp. *scabra*, *G. scabra* (Vis.) Chiov. ssp. *schimperi* (Sch. Bip. in Walp.) J. Baagøe, *G. villosa* Sch. Bip. In Walp., and *G. zavattarii* Lanza. However, based on a phylogenetic analysis of all species in the genus and other species in the tribe Helianthae, Geleta et al. (2010) suggested that the two subspecies of *G. scabra* should be treated as separate species. Additionally, the authors described two potential new species called "Chelelu" and "Ketcha" with their local names. However, their proper taxonomic description is yet to be completed in order to assign scientific names. The genus *Guizotia* will therefore have nine species when its taxonomy is revised in the future.

Each taxon is distinguishable from the other taxa by its growth pattern and morphological characteristics (Baagoee 1974). According to sequences derived from various chloroplast DNA regions, *G. abyssinica*, *G. scabra* ssp. scabra, *G. scabra* ssp. *schimperi* and *G. villosa* were found phylogenetically closely related. While *G. scabra* ssp. *schimperi* looked most closely related to *G. abyssinica* (Geleta *et al.* 2010). *G. abyssinica* (noug, also called niger) is an annual and cultivated taxon with clear distinguishable botanical attributes, such as its ovate outer phyllaries and large achene size. It is mainly cultivated in Ethiopia and India (Getinet & Sharma 1996).

2.2 The biology of noug

Noug is an erect, stout, branched, dicotyledonous annual herb, cultivated for its oilseeds (Getinet & Sharma 1996). It is closely related to sunflowers and differs from domesticated sunflowers mainly by its high level of branching, numerous flower heads, and small seeds (Dempewolf *et al.* 2015). It has epigeal germination with paired first leaves. Leaves are arranged on opposite sides of the stem, but at the top of the stem, leaves are arranged in an alternate fashion. Leaves are 10–20 cm long and 3–5 cm wide (Getinet & Sharma 1996). The leaf margin varies from coarse to smooth while the colour varies from light green to dark green with a smooth leaf surface (Getinet & Sharma 1996; Demissie *et al.* 2021). The stem is smooth to slightly rough and the plant is usually moderately to well branched. The noug stem is hollow and can be easily broken. The number of branches ranges from 5 to 12 and in a very dense plant stand, fewer branches are formed. The colour of the stem varies from dark purple to light green and the stem is about 1.5 cm in diameter at the base (Getinet & Sharma 1996; Ranganatha *et al.* 2016). The plant height is on average 1.4 m and can vary considerably based on environmental factors, and can reach up to 2 m. The flower is yellow. The heads are 15–50 mm in diameter with 5–20 mm long ray florets. Each composite flower has ray and disk florets. The receptacle has a semi-spherical shape and is 1–2 cm in diameter and 0.5–0.8 cm long. The receptacle is surrounded by two rows of bracts. The capitulum consists of 6–8 fertile female ray florets with narrowly elliptic, obovate ovules. The stigma has two curled branches. The hermaphrodite disk florets, usually 40–60 per capitulum, are arranged in three

whorls. The disk florets are yellow or orange with yellow or dark anthers, and a densely hairy stigma. The achene is club-shaped, obovoid, and narrow (Getinet & Sharma 1996). The number of seeds per capitulum ranges between 39–97 with average of 65 seeds per capitulum (Demissie *et al.* 2021) The head produces about 39 to 97 in range seeds. Noug is a strictly self-incompatible oilseed crop with a sporophytic self-incompatibility mechanism. This feature prevents self-pollination and encourages outcrossing (Geleta & Bryngelsson 2010; Geleta & Ortiz 2013). This phenomenon has great significance to plant breeding such as crop yield, quality improvement, marker-assisted breeding through self-incompatible genotyping, and development of hybrids for overcoming intra- and interspecific reproductive barriers (Muñoz-Sanz *et al.* 2020; Dwivedi 2021).

2.3 Origin, diversity center, domestication, and geographic distribution of noug

While archeobotanical evidence regarding the origin and domestication of noug is lacking, morphological, phytogeographical, and cytological evidence suggests that it was domesticated in Ethiopia as early as 3,000 BC from *G. scabra schimperi* through selecting plants with larger seeds and further cultivation (Harlan 1969; Baagøe, 1974; Hiremath & Murthy, 1988; Murthy, Dagne, 2001). However, the magnitude of genetic diversity and the diversity of its cultural and use values suggest that domestication might have occurred earlier than the time indicated here (Geleta 2007). A long history of cultivation suggests that Ethiopia is the center of origin, domestication, and diversity of *G. abyssinica* (Baagøe, 1974; Weiss, 1983; Hiremath & Murthy, 1988). Through Ethiopian immigrants and/or trade routes, the crop reached India (Hiremath & Murthy 1988), the second country where noug is cultivated at a significant level. It is also cultivated in some areas in Sudan, Uganda, Zaire, Tanzania, Malawi and Zimbabwe, and the West Indies, Nepal, Bangladesh, Bhutan and India (Weiss 1983). Baagoe (Baagoee 1974) indicated that noug was derived from *G. scabra* spp. *schimperi* and its specific characteristics were developed through cultivation. The analysis of the cytological and morphological data supports the suggestion that noug is derived from *G. scabra* ssp *schimperi* (Baagoee 1974). This is further supported

by DNA marker-based research that revealed *G. scabra* spp. *schimperi* is the most closely related taxon to noug (Geleta *et al.* 2007b; Geleta *et al.* 2007a; Geleta *et al.* 2010). As *G. arborescens*, *G. reptans*, and *G. zavattarii* differ in their morphology, habit, and geographic distribution from noug, none of these taxa could be the progenitors of noug.

2.4 Economical, nutritional, medicinal, cultural, and ecological significance of noug

Although noug is a minor oilseed crop globally, it is an economically significant crop serving as a major source of edible oil in Ethiopia. In India, it accounts for about 2% of its total oilseed production (Riley & Belayneh 1989; Osman 2020). It is also cultivated to a lesser extent in Sudan, Uganda, Tanzania, Malawi, and Zimbabwe (Vaughan 1970). Noug seed contains 30–50% oil (Seegeler 1983; Geleta, Stymne & Bryngelsson 2011; Tsehay *et al.* 2021). Its oil is mainly used for cooking (Geleta & Ortiz 2013; Tesfaye 2019). The oil is also used as component of soap, paints, and lubricants (Ramadan 2012). Noug press cake contains 33 to 37% protein (Baagoee 1974), which is used as animal feed, manure, and fuel. Noug seeds are prepared in different ways for human consumption. After roasting and grinding, it can be prepared into a cake and eaten mixed with roasted grains or as soup after mixing with water and sugar. The seeds are also used for fortification of diets and are sources of various compounds of nutritional value and positive health effects (Lima *et al.* 2021; Elbermawi *et al.* 2022; Panda *et al.* 2023). The crop is highly valued culturally in Ethiopia and India and as a medicinal plant (Bani, V & B 2023).It also has ecological significance, as it is a suitable crop as a cover crop (Büchi *et al.* 2020).

2.5 The noug genome

The number of chromosomes and genome size are key cytological characteristics of crop species reflecting their phylogenetic relationships with their wild and weedy relatives. Noug is a diploid species with a 2n chromosome number of 30) with 30 chromosomes (Murthy, Hiremath & Salimath 1993).

Cytological analysis revealed that all *Guizotia* species have the same chromosome number ($2n = 30$). However, there are significant differences in genome size and karyotypes. *G. abyssinica*, *G. scabra* ssp. *schimperi* and *G. villosa* have symmetrical karyotypes, whereas *G. scabra* ssp. *scabra*, *G. reptans* and *G. zavattarii* exhibited asymmetrical karyotypes (Hiremath & Murthy 1992; Hiremath, Murthy & Salimath 1992; Dagne 1995).The genome size of *Guizotia* species which is refereed as "C-value" is half of the amount of DNA in the somatic nucleus (Greilhuber *et al.* 2005). There is a difference in the nuclear DNA amount among Guizotia species which was quantified by (Hiremath, Murthy & Salimath 1992) using the method described in Ohri-Khoshoo and Hiremath-Salimath (Ohri & Khoshoo 1987; Hiremath & Salimath 1991). The study of this research back in 1992 revealed that there is no significant intraspecific DNA variation in both wild and cultivated species. Among the species of *Guizotia*, interspecific DNA genome size variation of about three-fold was recorded between 3.61 picogram for *G. reptans* and 11.37 picogram for *G. zavattarii* (Table 1). Here it is important to note that genome size can vary widely among different species, even within the same genus. Such information can be valuable for taxonomic and genetic studies, as well as for understanding the evolution and diversity of plants within the genus *Guizotia*.

Table 1. Average 2C value of nuclear DNA amount in Guizotia species.

Taxon	**Life form**	**Average 2C DNA amount in picogram**	**Chromosome number**	
			n	2n
G. abyssinica	Annual	7.57	15	30
G. schimperi	Annual	4.25	15	30
G. scabra	Perennial	5.26	15	30
G. villosa	Annual	4.21	15	30
G. reptans	Perennial	3.61	15	30
G. zavattarii	Perennial	11.37	15	30
G. arborescens	Perennial	Not known	15	30

The observation of an extra DNA accumulation of about 78% in *Guizotia abyssinica* compared to its wild progenitor *G. schimperi* suggests significant changes in the DNA content during the domestication and cultivation of *Gui-*

zotia abyssinica, which might be attributed to a combination of natural selection and human cultivation practices (Hiremath, Murthy & Salimath 1992; Murthy, Hiremath & Salimath 1993).

2.6 Molecular markers and their applications

Molecular markers are indeed genetic loci (specific positions in the DNA sequence) that can be easily tracked, quantified in a population and serve as valuable tools in genetic and genomic research (Dekkers & Hospital 2002; Kumar & Jorben 2020; Amiteye 2021). Molecular markers speed up breeding by facilitating genetic gains and shortening breeding cycles (Rajaguru, Annapurna & Rajeev K 2020; Singh *et al.* 2022). Morphological markers have limitations due to the fact that they can be affected by environmental factors and plant growth stages, are limited in number, and can be inefficient at delineating genetic variability (Nadeem *et al.* 2018). These limitations of morphological markers have necessitated the use of DNA-based markers for genetic diversity characterization and utilization in advanced plant breeding. In the case of noug, previous research used molecular markers for phylogenetic and genetic diversity analyses. Table 2 summarizes previous utilization of molecular markers in noug research.

Table 2. Molecular markers used in noug previous research.

Markers[z]	Objectives of the study	References
DNA/ITS	Phylogenetic analysis	(Bekele *et al.* 2007)
RAPD	Genetic diversity analysis	(Geleta *et al.* 2007c)
ISSR	Genetic diversity analysis	(Petros, Merker & Zeleke 2007)
RAPD	Genetic diversity analysis	(Nagella *et al.* 2008)
AFLP	Genetic diversity analysis	(Geleta *et al.* 2008)
cpDNA	Phylogenetic analysis	(Geleta *et al.* 2010)
EST-SSR	Genomic tool development	(Dempewolf *et al.* 2010)
ISSR	Genetic diversity analysis	(Hussain *et al.* 2015)
SSR	Domestication, genetic diversity	(Dempewolf *et al.* 2015)
SSR	Genetic diversity analysis	(Abebaw & Solomon 2017)
SSR	Genetic diversity analysis	(Mengistu, Abtew & Bitima 2019)
SSR	Genetic diversity analysis	(Mengistu, Abtew & Bitima 2019)
miRNAs	Abiotic stress	(Naik & Varadahalli 2020)
miRNAs	Abiotic stress	(Prathiba 2020)
ISSR	Genetic diversity analysis	(Moraa 2021)
SNP/KASP	Genomic tool development and population genetic analysis	(Tsehay *et al.* 2021)
SSR/SNP	Genomic tool development and genetic diversity analysis	(Gebeyehu *et al.* 2022)
SSR	genetic diversity and population structure analysis	(Terefe *et al.* 2023)

[z] DNA/ITS = DNA Internal Transcribed Spacer , RAPD = Random Amplified Polymorphic DNA, ISSR = Inter-Simple Sequence Repeat, AFLP = Amplified Fragment Length Polymorphism, cpDNA = Chloroplast DNA, EST-SSR = Expressed Sequence Tag-Simple Sequence Repeat, SSR = Simple Sequence Repeat, miRNAs = microRNAs, SNP/KASP = Single Nucleotide Polymorphism/Kompetitive Allele-Specific PCR.

As summarized above in Table 2, different types of molecular markers have been utilized mainly to understand the levels of genetic diversity, population structure, and phylogenetic relationship between noug and its weedy and wild relatives. However, the use of molecular markers on noug in other areas of research such as genome-wide association analysis (GWAS), bi-parental QTL mapping, and genomic selection is lacking.

2.7 Genotype by environment interactions (GEI) and methods for its estimation

Developing stable high-yielding cultivars with desired quality and host plant resistance to pathogens and pests are major goals of plant breeding (Shojaei *et al.* 2021; Tanin *et al.* 2022). The phenotypic performance of a cultivar or landrace population with regard to a specific trait is a cumulative result of its genotypic makeup, environment, and the interaction between genotype and environment (Pham *et al.* 2022). Genotype by environment interaction (GEI) is a dominant factor that challenges plant breeders in crop improvement (Hailemariam Habtegebriel 2022; Rebollo *et al.* 2023). Similar to the case in other crops, there is an increasing trend for GEI analysis in oilseed crops. Understanding GEI is required to develop resilient cultivars with low sensitivity to environmental factors in target cultivation areas (Malosetti, Ribaut & van Eeuwijk 2013), and a deep understanding of GEI facilitates a path to precision agriculture.

Different genotypes may respond differently to environmental factors. There are different scenarios with regard to genotype performance in different environments (Malosetti, Ribaut & van Eeuwijk 2013) as illustrated in Figure 2.

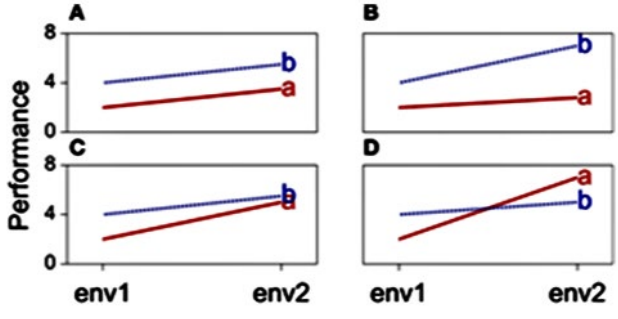


Figure 2. Genotype-by-environment interaction in terms of changing mean performances across environments: (A) additive model, (B) divergence, (C) convergence, (D) cross-over interaction.

Different approaches could be used to estimate GEI and genotype stability. The methods or models can generally be divided into two categories based

on whether they are parametric or non-parametric (Pour-Aboughadareh *et al.* 2022). Parametric models make assumptions about the underlying data distribution, which is not the case with non-parametric models. Parametric models assume that the data follows a particular distribution, most commonly a normal distribution (Shahbazi 2019). For example, the Additive Main Effects and Multiplicative Interaction (AMMI) model and the Genotype and Genotype by Environment (GGE) model assume normality in the data.

Non-parametric methods are often preferred when the assumptions of parametric models are not met, and the data distribution is not well behaved (Whitley & Ball 2002). They provide a more robust way to assess GEI and genotype stability. However, parametric methods may still be useful when data meet the distributional assumptions and are easier to interpret due to the explicit mathematical modeling of GEI. Non-parametric models may not provide as much insight into the underlying genetic and environmental effects as parametric models do. They are more focused on ranking and may not be suitable for estimating means, variances, or interaction effects.The choice between these two approaches should be made based on the nature of the data, research objectives, and the level of detail required. Researchers often use a combination of both parametric and non-parametric methods to gain a comprehensive understanding of GEI and genotype stability in plant breeding and agricultural research. A non-parametric model results in the reduction of the bias caused by outliers, and no assumptions are needed about the distribution of observed values when using this method (Sadiyah & Hadi 2023). They are more flexible and can handle data that may not fit a normal distribution (Huehn 1990; Abdipour *et al.* 2017).

2.8 Biotic stresses in noug

Biotic stresses can have a negative impact on the productivity of various crops, including oilseed crops like noug. While noug may have relatively better tolerance to biotic stresses compared to some other crops, it is still important to be aware of these potential stressors and implement preventive measures (Clements, DiTommaso & Hyvönen 2014; Goyal *et al.* 2022; Seethapathy, Gurudevan & Prabakar 2023). Various form of biotic stresses

such as insect pests, pathogens, pests and weeds affect noug cultivation (Getinet & Sharma 1996; Goyal *et al.* 2022). Research on biotic stresses is much less than that of many other crops of economic and nutritional significance, although there were previous scientific reports (Terefe *et al.* 2008). Records of noug diseases, insect pests and weeds can be found in (Getinet & Sharma 1996; Goyal *et al.* 2022).The most frequently observed diseases of noug include noug blight, powdery mildew, shot holes, and bacterial leaf spots (Getinet & Sharma 1996; Sujatha *et al.* 2023). Among noug diseases, shot hole causes about 10% yield loss as estimated based on the results of a fungicide screening trial. However, publications are unavailable regarding yield loss due to other diseases. Shot hole disease (Figure 3 A and B), as observed and identified by a plant pathologist during our field experiment in Holeta, is a widespread disease in noug growing areas in Ethiopia with varying intensity. Varying level of shot hole intensity was recorded in Amhara and Oromia regions of Ethiopia. As recorded at Holeta, the disease may cause up to 10% yield loss under severe infection conditions.

According to (Medhin & Mulatu 1992) several different species of insect pests of noug were reported . Among them, noug flies, black pollen beetles, and thrips are among the major ones that need research priorities in Ethiopia. African bollworm (Figure 3 D) is among noug insect pests but is not a major one. The very recognized weed of noug is dodder (*Cuscuta campestris*) which is parasitic and affects noug production though out Ethiopia (Getinet & Sharma 1996; Goyal *et al.* 2022). Integrated pest management (IPM) and sustainable farming practices, such as crop rotation, the use of resistant crop varieties, and cultural control methods, can play a crucial role in minimizing the impact of these biotic stresses on noug. Some of the main insect pests, pathogens, and weeds commonly found in Ethiopian agricultural settings are listed below in Table 3.

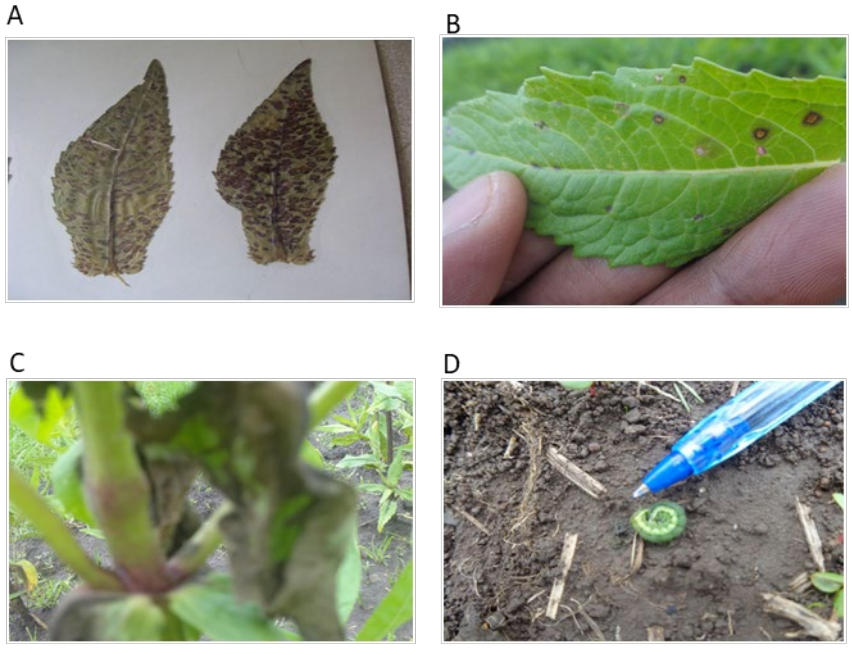


Figure 3. Diseases and an insect pest of noug observed during field experiments in Ethiopia: (A) Shot hole disease on mature leaf, (B) Shot hole disease on green leaf, (C) Leaf blight, and (D) African bollworm. Photo: Sewalem Tsehay (2017).

Table 3. Major forms of biotic stresses in noug.

Name	Biotic stresses	Damage	References
Insect Pests	*Eutretosoma* sp.	Feeds on the flowers	(Medhin & Mulatu 1992)
Insect Pest	*Meligethes* species (Black pollen beetle)	Flower heads	(Getinet & Sharma 1996; Goyal *et al.* 2022)
Insect Pest	*Haplothrips articulosus* (Noug flower thrips)	Flowers and on leaves	(Schmutterer 2009)
Insect Pest	*Helicoverpa armigera*	Damage heads and developing seeds	(Getinet & Sharma 1996)
Pests	Birds and rodents	Seeds	Field observation
Pathogen	*Alternaria* sp.	Leaves	(Goyal *et al.* 2022)
Pathogen	*Sphaerotheca* species	Leaves	(Sujatha *et al.* 2023)
Pathogen	*Septoria* species	Leaves	(Getinet & Sharma 1996; Gupta, Bisen & Tiwari 2018; Goyal *et al.* 2022)
Pathogen	*Alternaria porri* species (dauci)	Leaves	(Gupta, Bisen & Tiwari 2018; Goyal *et al.* 2022)
Pathogen	*Xanthomonas campestris*	Leaves	(Gupta, Bisen & Tiwari 2018)
Weed	*Cuscuta campestris*	Branches, flowers	(Getinet & Sharma 1996)

3. Aims and objectives of the thesis

3.1 General aims and objectives

This PhD thesis research was aimed at (1) assessing the nutritional profiles of noug genetic resources in Ethiopia thereby identifying genotypes with desirable nutritional profiles for use in the crop's breeding programs, (2) evaluating the effects of genotype-by-environment interactions on their oil content and fatty acid compositions thereby identifying stable genotypes with high nutritional values, as well as (3) developing genomic tools and resources for noug for various applications, such as efficient conservation of noug genetic resources as well as marker-trait-association analysis that facilitate genomics-driven noug breeding.

3.2 Specific aims and objectives

The specific aims and objectives were to:

- Investigate the quantity and quality of noug seed nutrients including oil content, fatty acids, proteins and minerals.
- Develop new genomic resources for various applications
- Discover single nucleotide polymorphisms (SNPs) and convert them into Kompetitive Allele Specific PCR (KASP) markers

- Genotyping noug using the newly developed SNP/KASP markers for population genetic analysis.
- Evaluate the effects of GEI on oil content and proportions of major fatty acids.

4. Materials and methods

4.1 Plant materials and experiments

Genetically diverse noug germplasm that includes landraces 28 collected from wide geographic areas in Ethiopia, 7 cultivars released in Ethiopia, and 10 breeding populations selected based on their oil and yield-related traits were used for nutritional profiling. For this, field trials were conducted in Ethiopia at three sites namely Holeta (09°04'N, 38°29'E; 2400 masl), Ginchi (9°01'N and 38°10'E; 2300 masl), and Debrezeit (8°44'N and 38°58'E; 1900 masl) from June to December 2016 (Figure 4). The field trials made it possible to determine the genotypic, environmental, and genotype-by-environment components of phenotypic variance in the targeted traits. For genomic tool development, two self-compatible breeding lines previously developed based on genotypes selected from landrace populations C19 and K13 (Geleta & Bryngelsson 2010) were used for transcriptome sequencing. These two lines were selected based on the significant variation between them in terms of traits, such as maturity time, seed size and shape, oil content, and fatty acid composition, to facilitate the identification of a large number of SNP markers. The identified SNP markers were converted to KASP markers and used to genotype 24 noug accessions. Thirty-one noug accessions, which were grown successfully at two experimental sites in Ethiopia (Holeta and Ginchi, Figure 4) for three years (2015, 2017, and 2018), were evaluated to study the genotype by environmental interaction.

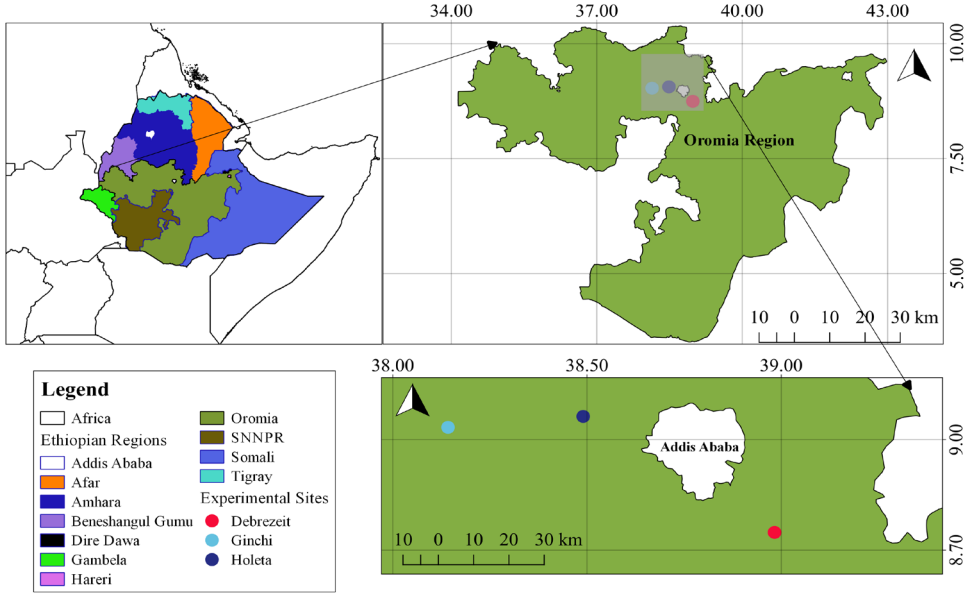


Figure 4. Field experimental sites in Ethiopia: Debrezeit, Ginchi and Holeta.

4.2 Methods

4.2.1 Lipid analysis

Several methods for lipid extraction from oilseeds have been developed, each with its advantages and disadvantages. The choice of method depends on factors like the type of oilseed, the desired quality of the extracted oil, and the available equipment (Kehelpannala *et al.* 2020). The most effective and cost-efficient method for extracting the oil is generally preferred. There are four basic methods of vegetable oil extraction: chemical, supercritical fluid, steam distillation, and mechanical (Sari 2006) as depicted in Figure 5.

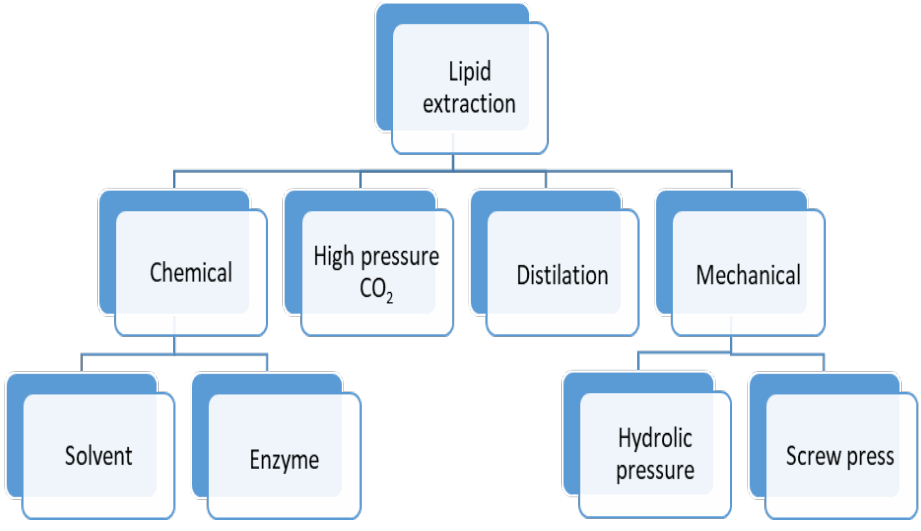


Figure 5. Basic methods for lipid extraction (Sari 2006).

In the research for this thesis, a solvent-based chemical oil extraction method was used. The details of the field trials, oil extraction method, and data analyses are provided in paper I of this thesis (Tsehay *et al.* 2021).

4.2.2 Determination of protein and mineral contents

Protein content was estimated based on the total nitrogen content of noug seeds analysed using the Kjeldahl method (Kjeldahl 1883). A nitrogen-to-protein conversion factor (NPCF) is used to determine protein content based on nitrogen content, which could be slightly different for different crops. Based on the review of (Krul 2019) on conversion factors, an NPCF of 5.6 was used in this study. The mineral contents in noug seeds were determined using a Perkin-Elmer Optima® 8300 inductively coupled with a plasma optical emission spectrometer (ICP-OES, Perkin-Elmer, Waltham, MA, USA) as described in paper I (Tsehay *et al.* 2021). Using this method, the levels of

12 mineral elements (Ca, Cu, Fe, K, Mg, Mn, Na, P, S, Se, Zn, and B) in 36 accessions were determined.

4.2.3 RNA and DNA extractions, transcriptome sequencing and assembly, and SNP calling

Total RNA was extracted from the leaves of two-week-old seedlings of two self-compatible genotypes using a spectrum plant total RNA kit from Sigma-Aldrich. To remove DNA from the extracted RNA, Qiagen RNase-free DNase (Qiagen, Stockach, Germany) was used. Library preparation, transcriptome sequencing, assembly, and SNP analysis were conducted at the Genome Sciences Center at the University of California, Davis. A non-normalized cDNA library was prepared following Illumina's guidelines and sequenced on the Illumina GAII according to (Hodgins *et al.* 2014), and then transcriptome assembly was done. A total of 4781 previously developed transcript sequences based on USDA-ARS accession PI 508077, which were functionally annotated using the *Arabidopsis* Information Resource (TAIR), were used as a reference for SNP discovery (Dempewolf *et al.* 2010). SNP calling was done after transcriptome assembly, and high-quality SNPs were obtained. A Detailed description of the method is given in paper II (Tsehay *et al.* 2020).

For DNA extraction, 24 noug accessions were planted in a greenhouse at the Swedish University of Agricultural Sciences (SLU) in Alnarp. DNA was extracted from the leaf tissue of two-week-old seedlings separately from individual plants for each accession using an LGC plant sample collection kit (KBS-9370-001) provided by LGC-Genomics. The extracted DNA was genotyped by LGC-Genomics following the conversion of SNP markers to KASP markers.

4.2.4 KASP assay design and genotyping

Transcriptome-based SNP markers were converted to KASP markers at LGC-Genomics. For this, 628 reference transcript sequences containing 959 SNP loci, including their alleles, were provided to LGC genomics. KASP assay was successfully designed for 931 SNPs. The 931 KASP markers were used for genotyping 281 genotypes from 24 noug accessions. The links to the details of the methods used for the analyses can be found in the paper II (Tsehay *et al.* 2020).

5. Results and discussion

5.1 Nutritional profiling

Oils and fats are among the main food items sources. Nowadays, these categories of nutrients are highly consumed replacing animal fats partially in processed food compositions (Sharma *et al.* 2022). Vegetable oils are a significant source of unsaturated fats and have become more prominent in processed foods (Mannucci *et al.* 2023). They are often used for frying, baking, and salad dressings. Noug is the main source of vegetable oil in Ethiopia (Geleta & Ortiz 2013).In this study, the seed oil content and fatty acid composition of noug seeds were determined using the gas chromatography bioanalytical method. The mean oil content of the analysed genotypes ranged between 33% and 46% with 15% of the genotypes having a mean oil content of 40% or above. The oil contents of noug have gradually increased during its domestication history (Riley & Belayneh 1989; Getinet & Sharma 1996). This could be due to the selection and domestication of noug genetic resources with high oil content in the seed over the years involving high oil genotypes. This was corroborated by research that revealed a population of noug up to 50% oil content in their seed (Geleta, Stymne & Bryngelsson 2011). Oilseed crops are significant sources of vegetable oil used for various purposes, including both food and industrial applications. These crops are cultivated primarily for the extraction of oil from their seeds. The choice of oilseed crop often depends on the specific characteristics and oil content of the seeds, as well as the desired end-use of the oil (Waseem *et al.* 2017).

Noug oil has high linoleic acid (72–79%) which is more similar to safflower (74–78%) and with total oil content in similar parity to other representative oilseed crops as described in the paper I (Tsehay *et al.* 2021). The intrinsic nature of noug considering its total oil in the seed is a great potential oilseed crop to improve it by breeding. Oleic (omega-9 fatty acid) and linoleic (omega-6 fatty acid) acids are both types of fatty acids in noug oil and they each have distinct health benefits (Alves *et al.* 2019; Tsehay *et al.* 2021). The health benefits of these acids include blood cholesterol, skin, heart, immune, anti-inflammatory properties, and weight management. Linoleic acid, which is the major component of noug oil and many other oil crops, is an essential polyunsaturated omega-6 fatty acid that the body cannot produce on its own (Jandacek 2017). Hence, noug can be one of the high-quality oilseed crops based on its attributes mentioned above. The noug oil content reported in paper I of this thesis (Tsehay *et al*. 2021) agreed with previous research (Seegeler 1983; Geleta, Stymne & Bryngelsson 2011), although higher oil content was recorded in some genotypes (Geleta, Stymne & Bryngelsson 2011).

Concerning fatty acid composition, the amount and proportions of fatty acids determine the nutritional features of vegetable oil (Chernova *et al.* 2019). Among the fatty acids found in noug seed oil, only the four major fatty acids were investigated in this thesis (paper I); namely, linoleic, oleic, palmitic, and stearic acids. Linoleic acid is the predominant fatty acid in noug seed oil ranging from 72–78%. Oleic acid, a monounsaturated fatty acid, was found in the range of 5–9%, while palmitic and stearic acids, both saturated fatty acids, ranged from 8 –10% and 7 –10%, respectively.

Similar results of fatty acid compositions were obtained in research of paper I and those of previous research on fatty acid composition of palmitic and stearic acids (Dutta *et al.* 1994; Dagne & Jonsson 1997; Yadav *et al.* 2012), while there was a difference in oleic and linoleic acids as observed previously (Dagne & Jonsson 1997; Yadav *et al.* 2012). Oleic acid was highly affected by environment and genotype-by-environment interactions relative to other fatty acids (Tsehay *et al.* 2021). The impacts of environment and genotype-by-environment interaction on oleic fatty acid were noticed in noug and other oilseed crops (Wolf et al. 1982; Lee et al. 2009; Geleta, Stymne &

Bryngelsson 2011; Sharma & Goyal 2015; Tomé-Rodríguez et al. 2023). The specific composition may vary depending on factors like the source, processing methods and growing conditions of the noug seeds. Noug oil also contains a small amount of other fatty acids namely palmitoleic, linolenic, arachidic, eicosenoic, behenic, erucic and lignoceric acids. These fatty acids make up about 2% to 3% of the total fatty acid content (Osman 2020).

When the whole seed or press cake is consumed as food or feed, noug serves as an important source of protein in addition to its vegetable oil. The present study revealed that noug seed protein content was in the range of 25–28%, a bit wider than the 26% range reported in a previous study (Thatte & Jyothi Lakshmi 2012). However, the analysis of variance (ANOVA) conducted on the protein data of the genotypes grown at two locations in Ethiopia (Holeta and Ginchi) revealed that genotypic variance was not significant (see Table 3 in paper I).

The mineral analysis of the noug seed showed limited genotypic variation for the 12 minerals analysed: Ca, Cu, Fe, K, Mg, Mn, Na, P, S, Se, Zn, and B. There was significant variation due to the experimental location (Tsehay *et al.* 2021), which might be due to soil plant nutrient content. This entails the consideration of environmental conditions, such as soil type and other biotic and abiotic factors, when we growing noug.

5.2 Development of genomic tools and analysis of noug population genetics

The major objectives of noug breeding include (1) increasing seed yield and seed oil content, and diversifying the seed oil fatty acid composition, and (2) developing cultivars that are self-compatible, pathogen and pest resistant, abiotic-stress tolerant, single-headed, shatter-proof, and non-lodging (Terefe & Girma 2022). Utilizing genetic variation in the number of heads per plant in noug genetic resources, accessions with the desired traits, could be crossbred to create offspring with similar traits by controlled pollination to

ensure the genetic material from the selected genotypes is passed on to the next generation (Getinet & Sharma 1996; Ranganatha *et al.* 2016). The use of genomic tools could improve the efficiency of noug improvement in these and other traits as well as for efficient conservation of its genetic resources. However, the genomic tools and resources previously developed for noug are insufficient for these purposes. Hence, concerted efforts to further develop such resources and tools are vital for noug breeding and conservation.

As provided in detail in paper II, new transcriptome-based SNP markers were developed and converted to bi-allelic KASP markers. These SNP/KASP markers were used in population genetic analysis as part of this thesis research (Tsehay *et al.* 2020). Nine hundred thirty-one SNPs were converted to KASP markers and used for genotyping 24 noug accessions. Among them, 554 KASP markers successfully genotyped the accessions, which is a 60% success rate. The reference sequences containing the SNPs used to develop KASP markers in this study are provided in Table S2 of Paper II. Not all single nucleotide polymorphism (SNP) sequences can be easily converted into KASP assays or other genotyping assays. Factors that can affect the successful implementation of KASP assays include paralogous sequences, proximity to other polymorphic sites, high GC content, secondary structures, allelic bias, sequence length, and the position of the target SNPs in the reference sequences (Cheon *et al.* 2020; Ur Rehman *et al.* 2021). To address these challenges, researchers may need to employ various strategies, such as optimizing primer design, adjusting PCR conditions, or even considering alternative genotyping methods when designing assays for challenging SNPs. Additionally, careful validation and quality control are essential to ensure the accuracy and reliability of genotyping results. About half of the genotyped loci were monomorphic indicating that the genetic variation between noug genotypes in the regions of the genome represented by the SNPs was relatively low. The effective number of alleles across the loci ranged from 1.43 to 1.48 with a mean of 1.45 suggesting a relatively low level of genetic diversity at these loci. The mean values for polymorphism information content (PIC), which is a measure of how informative a genetic marker is, and expected heterozygosity (He) of the loci were 0.24 and 0.29, respectively, with large variations in fixation indices. The mean value for gene flow (Nm) was 3.17. About 50% of the loci showed significant

deviation from Hardy-Weinberg equilibrium (HWE), suggesting that the loci are under the influence of natural or artificial selection or both. Loci deviating from Hardy-Weinberg equilibrium and large variations in fixation indices may suggest complex genetic dynamics and evolutionary processes (Abramovs, Brass & Tassabehji 2020) in the noug populations.

Different genetic diversity parameters were estimated for each ac-cession across the SNP loci including the percent polymorphic loci (PPL), Shannon diversity index (I), observed heterozygosity (Ho), expected heterozygosity (He), and fixation index (F) The mean values of these parameters were 78%, 0.40, 0.24, 0.26, and 0.08, respectively. The analysis of molecular variance (AMOVA) attributed 95.5% of the total variation to variation within accessions. Although the variation between accessions was only 4.5% of the total variation, it showed significant population differentiation ($FST = 0.045$, $p < 0.0001$). Principal coordinate analysis (PCoA) conducted to determine the relationship between the accessions revealed that the first two principal coordinates explained 44% of the total variation. According to the results of this analysis, most of the accessions were closely clustered together, suggesting a low degree of differentiation among them, in agreement with the findings of the AMOVA analysis. It was interesting to observe, however, that some accessions were clearly separated from the main cluster, as illustrated by the PCoA biplot in paper II (Figure 5). The STRUCTURE software was used to conduct an admixture model-based analysis to shed more light on noug population structure. Based on this analysis, three genetic populations were determined to be the optimal representative of the accessions used in the study.

The genetic diversity analysis indicates, in summary, that there is a high level of genetic diversity within individual accessions (95.5%), and there is a low degree of differentiation between accessions (4.5%). The PCoA results support low differentiation among most accessions, but some accessions show significant genetic differences. This information is valuable for understanding the genetic diversity and population structure of noug, which has strong implications for the conservation of noug genetic resources and utilization in its breeding programs.

5.3 Genotype by Environment Interactions (GEI)

The organisms' genetic potential is limited by the effects of their environments (Sadras, Rebetzke & Edmeades 2013; Boyce, Sokolowski & Robinson 2020; Yousaf *et al.* 2022). The genotype-by-environment interaction (GEI) changes genotype responses from environment to environment in certain traits or behaviours, and this potential limitation results in using different genotypes for cultivation in different environments (D'Aguillo *et al.* 2021; Begna 2022; Ligarreto–Moreno & Pimentel–Ladino 2022). Genotype interacts with its environment to determine its phenotype. Developing enhanced crops that can thrive in evolving climates, and withstand biotic and abiotic stress induced by global warming, is a pressing mandate within plant breeding initiatives (Ortiz 2023).

The research, based on diverse accessions grown across locations and years, presented in this thesis is probably the first on the GEI effects on noug traits, such as oil content and fatty acid composition. In this research, 31 noug accessions were investigated to determine the effects of GEI and genotype stability across environments. Additive main effect and multiplicative interaction (AMMI) and regression models were used. The GEI accounted for 25%, 57%, and 49% variation in total lipid, oleic and linoleic acids, respectively. Insights into GEI effects are necessary to evaluate crops' performance and stability under a changing climate. Hence, the evaluation of GEI by using diverse crop genetic resources is important for increasing the stability of desirable traits in cultivars released through breeding programs.

Given that noug is an oilseed crop, conservation and breeding are geared towards seed traits (Dempewolf *et al.* 2015). These traits are predominantly quantitative and show multi-genic inheritance, and consequently, environment and GEI could have significant effects on their phenotype, e.g. field performance (Ranganatha *et al.* 2016; Huang *et al.* 2020; Mittelsten Scheid 2022).

The present study used the AMMI model-based analysis of variance for 31 noug genotypes grown across six environments (Table 1 of paper III), as well as Finlay and Wilkinson modified joint regression analysis. Both revealed

highly significant difference for the interaction between genotypes, and testing environment as well as the sensitivity of genotypes to varying environmental conditions (Table 4).

Table 4. Analysis of variance based on Finlay and Wilkinson modified joint regression analysis.

Source	DF	SS	MS	VR	F pr.
Genotypes	30	2697.1	89.9	15.5	<0.001
Environments	5	5424.8	1084.9	187.6	<0.001
Sensitivities	30	1616.4	53.9	9.32	<0.001
Residual	306	1769.5	5.8		
Total	371	1107.8	31.0		

DF: degrees of freedom, SS: sum of squares, MS: mean square, VR: variance, F pr.: probability value.

The present study revealed that genotype, environment, and GEI had highly significant effects on variation in total lipid, oleic, and linoleic acids ($P < 0.001$). Hence, GEI could be exploited to select superior genotypes for specific target environments or more stable ones for cultivation across a wide range of environments (Zemour *et al.* 2021). On the other hand, GEI did not have significant effects on palmitic and stearic acids. The variations in these two traits were mainly due to genotypes although environments also had significant effects. Significant variation in trait stability have been shown in the key trait oil content based on the regression coefficients of the Eberhart-Russell and Finlay-Wilkinson analysis (Figure 6).

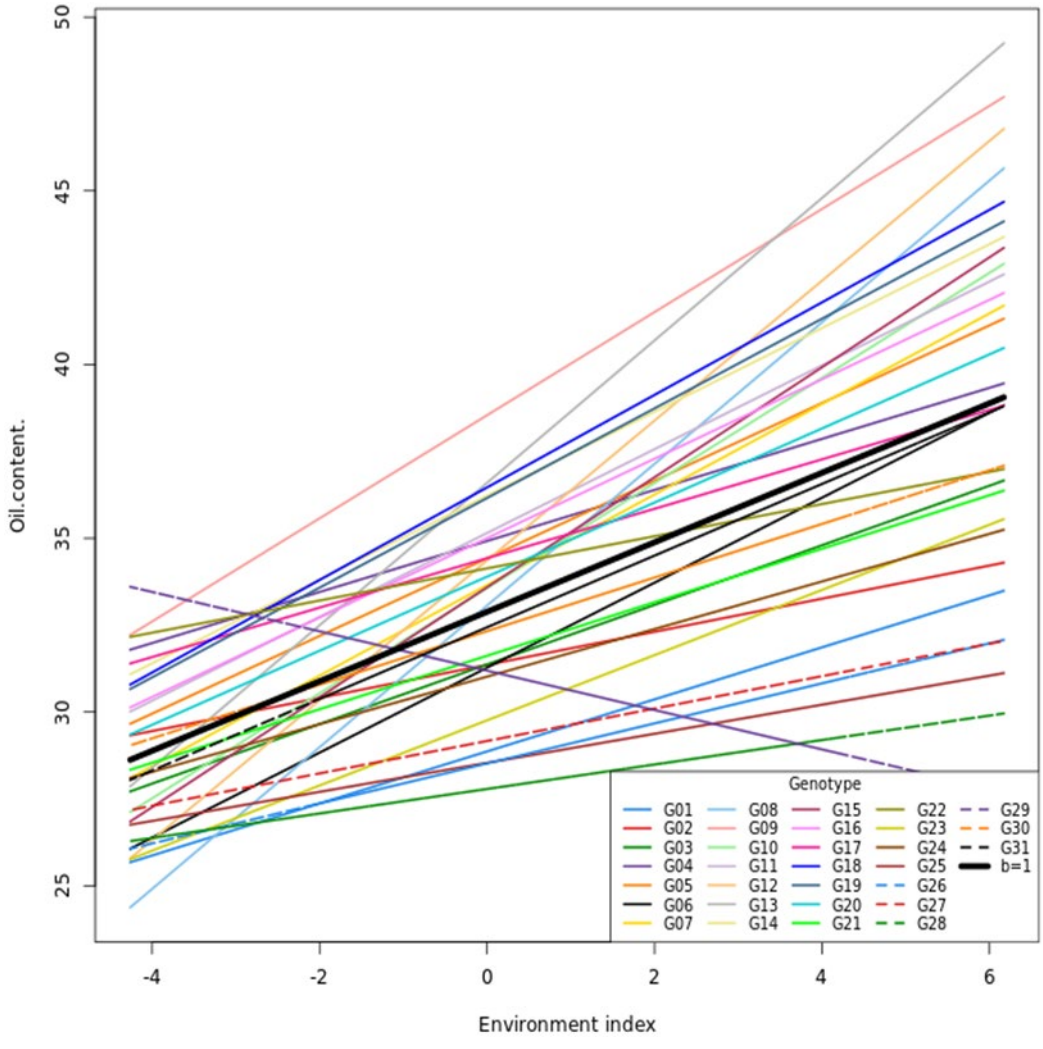


Figure 6. Reaction norm for oil content of 31 noug genotypes cultivated in 6 environments in Ethiopia.

The variation in the slope of these reaction norms directly related to GEI reflects the environmental sensitivity of the genotypes and shows how different genotypes will perform in different environments and can guide breeding or selection strategies to optimize desired traits, oil content, under specific conditions (Falconer 1990; Morrissey & Liefting 2016; Angeloni, Aguirrezábal & Echarte 2021).

6. Conclusions

The analysis of nutritional contents and determination of the amount and proportion of nutrients in a crop is crucial in determining its significance. Noug is among oil seed crops rich in lipids and fatty acid components, protein, and minerals. It is a highly suitable food crop, which can be consumed in the form of vegetable oil or processed whole seeds. Its oil is rich in linoleic acid (72-78%), an essential polyunsaturated fatty acid. The oleic acid content of some noug genotypes significantly increased when grown at low altitudes compared to their oleic acid content at higher altitudes. Such genotypes are suitable for cultivation at low altitudes to meet the high demands of vegetable oil rich in oleic acid, which is suitable for frying.

Based on transcriptome sequencing, novel SNP markers were developed for noug in this research. More than 50% of these SNPs were converted to KASP markers and utilized for population genetic analysis. These markers revealed a lower but significant population differentiation than other marker types previously used in noug. This could be attributed to the fact that they are transcriptome-derived and biallelic markers. In general, these newly developed markers will be useful for population genetic research, including genome-wide association analysis. Some of these markers may have significant associations with desirable traits in noug, thereby making them suitable for marker-aided breeding.

Developing highly productive cultivars across various environments is a key objective of plant breeding programs. Nevertheless, the GEI remains a major challenge to breed cultivars with stable performance across sites and over the years. The analyses of oil content and fatty acid composition data of noug genotypes grown in six environments using different statistical models (AMMI, Eberhart and Russell, Finlay-Wilkinson regression) revealed the significant effects of GEI on the crop's performance stability. The assessment of the GEI useful to identify traits that were significantly affected by GEI as well. Further, the stability analyses were useful to identify noug germplasm with high oil content and desirable fatty acid composition in a specific environment as well as across diverse environments.

7. Thesis research recommendations and future perspectives

In this thesis, the focus was on the investigation of nutritional profiling, the development of genomic resources for various applications, including population genetics and marker-trait association analyses, and the analyses of the effect of genotype-by-environment interaction on oil content and quality traits. Although the results obtained in this thesis are highly significant for the conservation and improvement of noug, they also indicate the need for further research. The following are major recommendations based on this thesis:

- In this research, a limited number of accessions grown in three environments were used for nutritional profiling. In order to have a deep insight into the nutritional values of noug genetic resources, a comprehensive profiling of noug nutrients involving a diverse germplasm representative of its noug gene pool, grown under various soil types and climates should be carried out. By using this approach, its nutritional values can be effectively improved.
- The analysis of seed protein compositions and quality should be carried out as part of a comprehensive nutritional profiling of noug genetic resources.

- This research has developed highly useful genomic resources. However, additional informative markers should be developed for noug in order to gain a deeper understanding of its genome through approaches such as genetic linkage mapping, which is currently unavailable.
- The markers developed in this research shall be utilized in marker-trait association analysis targeting major desirable traits in noug, preferably in combination with markers developed through further research. Markers significantly associated with traits of interest could then be used in marker-aided breeding. Moreover, the markers may be incorporated into a set of markers for future genomic selection-based noug breeding.
- It is highly desirable to perform next-generation sequencing-based transcriptome analysis (RNA-Seq) of noug genotypes with varying phenological, agronomic, and nutritional characteristics, as this provides highly informative novel markers and enables the identification of genes that regulate desirable traits in noug.
- Exploration of noug transcriptomes, proteomes, and metabolomes for deeper insights into the biological processes and mechanisms involved in the regulations of major traits, such as seed oil content and composition, thereby facilitating their improvements. A detailed analysis of how noug responds to different abiotic and biotic stresses, including the effects of climate change. Climate change exerts multifaceted impacts on crops, extending beyond seed yields to encompass critical quality attributes such as fatty acid composition and oil accumulation. Investigation into major oil biosynthesis pathways is vital in elucidating how prevalent forms of abiotic stress, particularly temperature fluctuations and water deficit, influence the quality and quantity of seed oil. A comprehensive understanding of the genotype-by-environment interaction, as it pertains to oilseed quality, is imperative for developing elite cultivars capable of generating high-quality seeds in the face of stress conditions induced by climatic changes.

- Develop inbred lines that possess a variety of desirable traits that are not adversely affected by inbreeding depression, and use them to produce F1 hybrids exhibiting hybrid vigor.
- To identify germplasm with stable seed yields and oil yields across diverse environments, further exploration of noug gene pool should be undertaken.

References

Abdipour, M., Vaezi, B., Younessi-Hamzekhanlu, M. & Ramazani, S.H.R. (2017). Nonparametric phenotypic stability analysis in advanced barley (*Hordeum vulgare* L.) genotypes. *Journal of Crop Science and Biotechnology,* 20(4), 305-314. https://doi.org/10.1007/s12892-017-0050-0

Abebaw, M. & Solomon, A. (2017). Genetic diversity assessment of *Guzoita abyssinica* using EST derived simple sequence repeats (SSRs) markers. *African Journal of Plant Science,* 11(4), 79-85. https://doi.org/10.5897/AJPS2016.1512

Abramovs, N., Brass, A. & Tassabehji, M. (2020). Hardy-Weinberg equilibrium in the large scale genomic sequencing era. *Frontiers in Genetics,* 11, 210. https://doi.org/10.3389/fgene.2020.00210

Agegnehu, G. (2010). Crop management practices in noug (*Guizotia abyssinica*) in the central highlands of Ethiopia. In: Geremew, T., Adugna, W. & Dereje, G. (eds) *The second oil crops research and development review workshop*, March 2010, Addis Ababa, Ethiopia. Ethiopian Institute of Agricultural Research. 240-245. https://www.researchgate.net/publication/239385406

Ahmad, M., Waraich, E.A., Skalicky, M., Hussain, S., Zulfiqar, U., Anjum, M.Z., Habib ur Rahman, M., Brestic, M., Ratnasekera, D., Lamilla-Tamayo, L., Al-Ashkar, I. & EL Sabagh, A. (2021). Adaptation strategies to improve the resistance of oilseed crops to heat Stress under a changing climate: an overview. *Frontiers in Plant Science,* 12, 767150. https://doi.org/10.3389/fpls.2021.767150

Alemaw, G. & Gurmu, F. (2023). Towards edible oil self-sufficiency in Ethiopia: lessons and prospects. *Cogent Food & Agriculture,* 9(1), 2198742. https://doi.org/10.1080/23311932.2023.2198742

Alves, A.Q., da Silva, V.A.J., Góes, A.J.S., Silva, M.S., de Oliveira, G.G., Bastos, I.V.G.A., de Castro Neto, A.G. & Alves, A.J. (2019). The fatty acid composition of vegetable oils and their potential use in wound care. *Advances in Skin & Wound Care,* 32(8), 1-8. https://doi.org/10.1097/01.Asw.0000557832.86268.64

Amiteye, S. (2021). Basic concepts and methodologies of DNA marker systems in plant molecular breeding. *Heliyon,* 7(10), 8093. https://doi.org/10.1016/j.heliyon.2021.e08093

Angeloni, P., Aguirrezábal, L. & Echarte, M.M. (2021). Assessing the mechanisms underlying sunflower grain weight and oil content responses to temperature during grain filling. *Field Crops Research,* 262, 108040. https://doi.org/https://doi.org/10.1016/j.fcr.2020.108040

Attia, Z., Pogoda, C.S., Reinert, S., Kane, N.C. & Hulke, B.S. (2021). Breeding for sustainable oilseed crop yield and quality in a changing climate. *Theoretical and Applied Genetics,* 134(6), 1817-1827. https://doi.org/10.1007/s00122-021-03770-w

Baagoee, J. (1974). The genus Guizotia (Compositae). A taxonomic revision. *Botanisk Tidsskrift,* 69, 1-39.

Bani, S., V, M. & B, K. (2023). Pharmacological and phytochemical profile of *Guizotia abyssinica* Cass – a review beyond its culinary utility. *RGUHS Journal of AYUSH Sciences,* 10(1), 14-20. https://doi.org/10.26463/rjas.10_1_7

Begna, T. (2022). Application of genotype by environmental interaction in crop plant enhancement. 8(2), 2454-6224. https://doi.org/10.20431/2454-6224.0802001

Bekele, E., Geleta, M., Dagne, K., Jones, A.L., Barnes, I., Bradman, N. & Thomas, M.G. (2007). Molecular phylogeny of genus *Guizotia* (Asteraceae) using DNA sequences derived from ITS. *Genetic Resources and Crop Evolution,* 54(7), 1419-1427. https://doi.org/10.1007/s10722-006-9126-0

Boyce, W.T., Sokolowski, M.B. & Robinson, G.E. (2020). Genes and environments, development and time. *Proceedings of the National Academy of Sciences,* 117(38), 23235-23241. https://doi.org/doi:10.1073/pnas.2016710117

Büchi, L., Wendling, M., Amossé, C., Jeangros, B. & Charles, R. (2020). Cover crops to secure weed control strategies in a maize crop with reduced tillage. *Field Crops Research,* 247, 107583. https://doi.org/https://doi.org/10.1016/j.fcr.2019.107583

Cheon, K.S., Jeong, Y.M., Oh, H., Oh, J., Kang, D.Y., Kim, N., Lee, E., Baek, J., Kim, S.L., Choi, I., Yoon, I.S., Kim, K.H., Won, Y.J., Cho, Y.I., Han, J.H. & Ji, H. (2020). Development of 454 new kompetitive allele-specific PCR (KASP) markers for temperate *japonica* rice varieties. *Plants,* 9(11), 1531. https://doi.org/10.3390/plants9111531

Chernova, A., Mazin, P., Goryunova, S., Goryunov, D., Demurin, Y., Gorlova, L., Vanyushkina, A., Mair, W., Anikanov, N., Yushina, E., Pavlova, A., Martynova, E., Garkusha, S., Mukhina, Z., Savenko, E. & Khaitovich, P. (2019). Ultra-performance liquid chromatography-mass spectrometry for precise fatty acid profiling of oilseed crops. *PeerJ,* 7, 6547. https://doi.org/10.7717/peerj.6547

Clements, D.R., DiTommaso, A. & Hyvönen, T. (2014). Ecology and Management of Weeds in a Changing Climate. In: Chauhan, B.S. & Mahajan, G. (eds) *Recent Advances in Weed Management*. New York, NY: Springer New York. 13-37. https://doi.org/10.1007/978-1-4939-1019-9_2

D'Aguillo, M., Hazelwood, C., Quarles, B. & Donohue, K. (2021). Genetic consequences of biologically altered environments. *Journal of Heredity,* 113(1), 26-36. https://doi.org/10.1093/jhered/esab047

Dagne, K. (1995). Karyotypes, C-banding and nucleolar numbers in *Guizotia* (Compositae). *Plant Systematics and Evolution,* 195(1), 121-135. https://doi.org/10.1007/BF00982319

Dagne, K. & Jonsson, A. (1997). Oil content and fatty acid composition of seeds of *Guizotia* Cass (Compositae). *Journal of the Science of Food and Agriculture,* 73(3), 274-278. https://doi.org/https://doi.org/10.1002/(SICI)1097-0010(199703)73:3<274::AID-JSFA725>3.0.CO;2-F

Dekkers, J.C.M. & Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics,* 3(1), 22-32. https://doi.org/10.1038/nrg701

Deme, T., Haki, G.D., Retta, N., Woldegiorgis, A. & Geleta, M. (2021). Fatty acid profile, total phenolic content, and antioxidant activity of niger seed (*Guizotia abyssinica*) and Linseed (*Linum usitatissimum*). *Front Nutr,* 8, 674882. https://doi.org/10.3389/fnut.2021.674882

Demissie, A., Hammenhag, C., Ortiz, R., Tesfaye, K. & Geleta, M. (2021). Characterization of oilseed crop noug (*Guizotia abyssinica*) using agro-morphological traits. *Agronomy,* 11, 1479. https://doi.org/10.3390/agronomy11081479

Dempewolf, H., Kane, N.C., Ostevik, K.L., Geleta, M., Barker, M.S., Lai, Z., Stewart, M.L., Bekele, E., Engels, J.M.M., Cronk, Q.C.B. & Rieseberg, L.H. (2010). Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.—the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Molecular Ecology Resources,* 10(6), 1048-1058. https://doi.org/https://doi.org/10.1111/j.1755-0998.2010.02859.x

Dempewolf, H., Tesfaye, M., Teshome, A., Bjorkman, A.D., Andrew, R.L., Scascitelli, M., Black, S., Bekele, E., Engels, J.M.M., Cronk, Q.C.B. & Rieseberg, L.H. (2015). Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). *Evolutionary Applications,* 8(5), 464-475. https://doi.org/https://doi.org/10.1111/eva.12256

Durrett, T.P., Benning, C. & Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. *The Plant Journal,* 54(4), 593-607. https://doi.org/https://doi.org/10.1111/j.1365-313X.2008.03442.x

Dutta, P.C., Helmersson, S., Kebedu, E., Alema, G. & Appelqvist, L.-Å. (1994). Variation in lipid composition of niger seed (*Guizotia abyssinica* Cass.) Samples collected from different regions in ethiopia. *Journal of the American Oil Chemists' Society,* 71(8), 839-843. https://doi.org/10.1007/BF02540459

Dwivedi, N. (2021). Self-incompatibility -a boon for plant breeders in hybrid seed production. *Agrospheres,* 2(2), 13-15.

Elbermawi, A., Darwish, M.S., Zaki, A.A., Abou-Zeid, N.A., Taher, M.A., Khojah, E., Bokhari, S.A. & Soliman, A.F. (2022). In vitro antidiabetic, antioxidant, and prebiotic activities of the chemical compounds isolated from *Guizotia abyssinica*. *Antioxidants* 11(12), 2482. https://doi.org/10.3390/antiox11122482

Falconer, D.S. (1990). Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genetics Research,* 56(1), 57-70. https://doi.org/10.1017/S0016672300028883

Frančáková, H., Ivanišová, E., Dráb, Š., Krajčovič, T., Tokár, M., Mareček, J. & Musilová, J. (2015). Composition of fatty acids in selected vegetable oils. *Potravinarstvo,* 9, 538-542. https://doi.org/10.5219/556

Gebeyehu, A., Hammenhag, C., Tesfaye, K., Vetukuri, R.R., Ortiz, R. & Geleta, M. (2022). RNA-seq provides novel genomic resources for noug (*Guizotia abyssinica*) and reveals microsatellite frequency and distribution in its transcriptome. *Front Plant Sci,* 13, 882136. https://doi.org/10.3389/fpls.2022.882136

Geleta, M. (2007). *Genetic diversity, phylogenetics and molecular systematics of Guizotia Cass. (Asteraceae). Diss.* Department of Plant Protection Biology Alarp: Swedish University of Agricultural Sciences. https://res.slu.se/id/publ/14267

Geleta, M., Bekele, E., Dagne, K. & Bryngelsson, T. (2010). Phylogenetics and taxonomic delimitation of the genus *Guizotia* (Asteraceae) based on sequences derived from various chloroplast DNA regions. *Plant Systematics and Evolution,* 289(1), 77-89. https://doi.org/10.1007/s00606-010-0334-x

Geleta, M. & Bryngelsson, T. (2010). Population genetics of self-incompatibility and developing self-compatible genotypes in niger (*Guizotia abyssinica*). *Euphytica,* 176(3), 417-430. https://doi.org/10.1007/s10681-010-0184-1

Geleta, M., Bryngelsson, T., Bekele, E. & Dagne, K. (2007a). Comparative analysis of genetic relationship and diagnostic markers of several taxa of *Guizotia* Cass. (Asteraceae) as revealed by AFLPs and RAPDs. *Plant Syst Evol.,* 265, 221-239. https://doi.org/10.1007/s00606-007-0521-6

Geleta, M., Bryngelsson, T., Bekele, E. & Dagne, K. (2007b). AFLP and RAPD analyses of genetic diversity of wild and/or weedy *Guizotia* (Asteraceae) from Ethiopia. *Hereditas,* 144(2), 53-62. https://doi.org/https://doi.org/10.1111/j.2007.0018-0661.01983.x

Geleta, M., Bryngelsson, T., Bekele, E. & Dagne, K. (2007c). Genetic diversity of *Guizotia* abyssinica (L. f.) Cass. (Asteraceae) from Ethiopia as revealed by random amplified polymorphic DNA (RAPD). *Genetic Resources and Crop Evolution,* 54(3), 601-614. https://doi.org/10.1007/s10722-006-0018-0

Geleta, M., Bryngelsson, T., Bekele, E. & Dagne, K. (2008). Assessment of genetic diversity of *Guizotia abyssinica* (L.f.) Cass. (Asteraceae) from Ethiopia

using amplified fragment length polymorphism. *Plant Genetic Resources,* 6(1), 41-51. https://doi.org/10.1017/S1479262108913903

Geleta, M. & Ortiz, R. (2013). The importance of *Guizotia abyssinica* (niger) for sustainable food security in Ethiopia. *Genetic Resources and Crop Evolution,* 60(5), 1763-1770. https://doi.org/10.1007/s10722-013-9997-9

Geleta, M., Stymne, S. & Bryngelsson, T. (2011). Variation and inheritance of oil content and fatty acid composition in niger (*Guizotia abyssinica*). *Journal of Food Composition and Analysis,* 24(7), 995-1003. https://doi.org/https://doi.org/10.1016/j.jfca.2010.12.010

Genetics, N. (2020). Crop genomes and beyond. *Nature Genetics,* 52, 865-865. https://doi.org/10.1038/s41588-020-0691-5

Getinet, A. & Sharma, S.M. (1996). *Niger. Guizotia abyssinica (L.f.) Cass. Promoting the conservation and use of underutilized and neglected crops. 5., IPK and IPGRI*. Gatersleben: International Plant Genetic Resources Institut. https://www.researchgate.net/publication/243055904

Giberti, G.C. (2018). Ancient and modern concepts about the asteraceae taxonomy. In: Sülsen & Martino (eds) *Sesquiterpene Lactones*. 1 edition. Springer Cham. 19-29. https://doi.org/https://doi.org/10.1007/978-3-319-78274-4_2

Goyal, V.K., Agrawal, N., Jain, M., Bisen, R. & Tripathi, R. (2022). Niger (*Guizotia abyssinica* (L. f.) Cass.) an oilseed crop under biotic stress. In: Hasanuzzaman, Ahammed & Nahar (eds) *Managing Plant Production Under Changing Environment*. 1 edition. Springer Singapore. 135-149. https://doi.org/10.1007/978-981-16-5059-8_6

Greilhuber, J., Doležel, J.L., A., M. & Bennett, M.D. (2005). The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-Value' to describe nuclear DNA contents. *Annals of Botany,* 95(1), 255-260. https://doi.org/10.1093/aob/mci019

Gunstone, F.D. (2011). Production and trade of vegetable oils. In: Gunstone (ed.) *Vegetable Oils in Food Technology: Composition, Properties and Uses*. 2 edition. Blackwell. 1-24. https://doi.org/https://doi.org/10.1002/9781444339925.ch1

Gupta, K., Bisen, R. & Tiwari, A. (2018). A review: current status of niger diseases and their integrated management. *International Journal of Chemical Studies,* 6(6), 2131-2135.

Hailemariam Habtegebriel, M. (2022). Adaptability and stability for soybean yield by AMMI and GGE models in Ethiopia. *Frontiers in Plant Science,* 13, 950992. https://doi.org/10.3389/fpls.2022.950992

Haroon, M., Mubashar Zafar, M., Farooq, M., Afzal, R., Batool, M., Idrees, F., Babar, U. & Saboor Khan, A. (2020). Novel plant breeding techniques shake hands with cereals to increase production. *Plants* 11, 1052. https://doi.org/https://doi.org/10.3390/plants11081052

Hiremath, S.C. & Murthy, H.N. (1988). Domestication of niger (*Guizotia abyssinica*). *Euphytica,* 37, 225-228.

Hiremath, S.C. & Murthy, H.N. (1992). Cytogenetical studies in *Guizotia* (Asteraceae). *Caryologia,* 45, 69-82.
Hiremath, S.C., Murthy, H.N. & Salimath, S.S. (1992). Quantitative nuclear DNA differences associated with genome evolution in *Guizotia* (Compositae). *Genetica,* 85(3), 241-247. https://doi.org/10.1007/BF00132276
Hiremath, S.C. & Salimath, S.S. (1991). Quantitative nuclear DNA changes in *Eleusine* (Gramineae). *Plant Systematics and Evolution,* 178(3), 225-233. https://doi.org/10.1007/BF00937965
Hodgins, K.A., Lai, Z., Oliveira, L.O., Still, D.W., Scascitelli, M., Barker, M.S., Kane, N.C., Dempewolf, H., Kozik, A., Kesseli, R.V., Burke, J.M., Michelmore, R.W. & Rieseberg, L.H. (2014). Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Mol Ecol Resour,* 14(1), 166-77. https://doi.org/10.1111/1755-0998.12163
Huang, W., Carbone, M.A., Lyman, R.F., Anholt, R.R.H. & Mackay, T.F.C. (2020). Genotype by environment interaction for gene expression in Drosophila melanogaster. *Nature Communications,* 11(1), 5451. https://doi.org/10.1038/s41467-020-19131-y
Huehn, M. (1990). Nonparametric measures of phenotypic stability *Euphytica,* 47(3), 189-194. https://doi.org/10.1007/BF00024241
Hussain, Z., Yadav, S., Kumar, S., Suneja, P., Nizar, M., Yadav, S., Yadav, S.K. & Dutta, M. (2015). Molecular characterization of niger [*Guizotia abyssinica* (Lf) Cass.] germplasms diverse for oil parameters. *Indian Journal of Biotechnology,* 14(3), 344-350.
Jandacek, R.J. (2017). Linoleic Acid: A nutritional quandary. *Healthcare* 5(2), 25. https://doi.org/10.3390/healthcare5020025
Kehelpannala, C., Rupasinghe, T.W.T., Hennessy, T., Bradley, D., Ebert, B. & Roessner, U. (2020). A comprehensive comparison of four methods for extracting lipids from *Arabidopsis* tissues. *Plant Methods,* 16(1), 155. https://doi.org/10.1186/s13007-020-00697-z
Kjeldahl, J. (1883). Neue methode zur bestimmung des stickstoffs in organischen körpern. *Zeitschrift für Analytische Chemie,* 22(1), 366-382. https://doi.org/10.1007/BF01338151
Krul, E.S. (2019). Calculation of nitrogen-to-protein conversion factors:a review with a focus on soy protein. *Journal of the American Oil Chemists' Society,* 96(4), 339-364. https://doi.org/https://doi.org/10.1002/aocs.12196
Kumar, A., Sharma, A. & Upadhyaya, K.C. (2016). Vegetable oil: nutritional and industrial perspective. *Curr Genomics,* 17(3), 230-40. https://doi.org/10.2174/1389202917666160202220107
Kumar, R. & Jorben, J. (2020). Molecular Markers: principles and methodology. In: Pidigam, S. (ed.) *Advances in Genetics and Plant Breeding*. (10). 89-106. https://doi.org/10.22271/ed.book.935
Lee, J.-D., Woolard, M., Sleper, D., Smith, R., Pantalone, V., Nyinyi, C., Cardinal, A. & Shannon, J. (2009). Environmental effects on oleic acid in soybean

seed oil of plant introductions with elevated oleic concentration. *Crop Science,* 49, 762–1768. https://doi.org/10.2135/cropsci2008.11.0663

Lee, K.-R., Kim, E.-H., Kim, K.-H., Park, J.-S. & Kim, H. (2017). Vegetable oil production in vegetative plant tissues. *Plant Biotechnology Reports,* 11, 385–395. https://doi.org/10.1007/s11816-017-0460-9

Li, J., Yang, Y., Sun, X., Liu, R., Xia, W., Shi, P., Zhou, L., Wang, Y., Wu, Y., Lei, X. & Xiao, Y. (2022). Development of intron polymorphism markers and their association with fatty acid component variation in oil palm. *Frontiers in Plant Science,* 13, 885418. https://doi.org/10.3389/fpls.2022.885418

Ligarreto–Moreno, G. & Pimentel–Ladino, C. (2022). Grain yield and genotype x environment interaction in bean cultivars with different growth habits. *Plant Production Science,* 25(2), 232-241. https://doi.org/10.1080/1343943X.2021.1981141

Lima, J.F., Días, M.I., Pereira, C., Ivanov, M., Soković, M.D., Steinmacher, N.C., Ferreira, I.C.F.R. & Barros, L. (2021). Characterization of nonconventional food plants seeds *Guizotia abyssinica* (L.f.) Cass., *Panicum miliaceum* L., and *Phalaris canariensis* L. for application in the bakery industry. *Agronomy,* 11(9), 1873. https://doi.org/https://doi.org/10.3390/agronomy11091873

Malosetti, M., Ribaut, J.-M. & van Eeuwijk, F.A. (2013). The statistical analysis of multi-environment data: modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology,* 4, 44. https://doi.org/10.3389/fphys.2013.00044

Mannucci, P.M., Jolliet, O., Meijaard, E., Slavin, J., Rasetti, M., Aleta, A., Moreno, Y. & Agostoni, C. (2023). Sustainable nutrition and the case of vegetable oils to match present and future dietary needs. *Frontiers in Public Health,* 11, 1106083. https://doi.org/10.3389/fpubh.2023.1106083

Medhin, T.G. & Mulatu, B. (1992). Insect pests of noug, linseed and *Brassica*. In: *Proceedings of the First National Oilseed Workshop*, 3-5 December, Addis Ababa 174-177.

Mengistu, B., Abtew, W.G. & Bitima, T.D. (2019). Estimating the genetic diversity of Ethiopian noug (*Guizotia abyssinica* (L.f.) Cass.) genotypes using SSR markers. *Adv Crop Sci Tech,* 7(2), 1000433. https://doi.org/http://dx.doi.org/10.4172/2329-8863.1000433

Mir, R.R., Kudapa, H., Pramod, S. & Lewis, R.S. (2022). Editorial: Biotechnological and genomic approaches for enhancing agronomic performance of crops. *Frontiers in Genetics,* 13, 991630. https://doi.org/10.3389/fgene.2022.991630

Mittelsten Scheid, O. (2022). Mendelian and non-Mendelian genetics in model plants. *Plant Cell,* 34(7), 2455-2461. https://doi.org/10.1093/plcell/koac070

Mohd Saad, N.S., Severn-Ellis, A.A., Pradhan, A., Edwards, D. & Batley, J. (2021). Genomics armed with diversity leads the way in *Brassica*

improvement in a changing global environment. *Front Genet,* 12, 600789. https://doi.org/10.3389/fgene.2021.600789

Moraa, O.L. (2021). Assessment of genetic diversity of niger plant (*Guizotia abyssinica* L.) in Moiben Sub County, Kenya, using inter simple sequence repeat markers. *International Journal for Research in Applied Sciences and Biotechnology,* 8(1), 126-131.

Morrissey, M.B. & Liefting, M. (2016). Variation in reaction norms: statistical considerations and biological interpretation. *Evolution,* 70(9), 1944-59. https://doi.org/10.1111/evo.13003

Msanne, J., Kim, H. & Cahoon, E.B. (2020). Biotechnology tools and applications for development of oilseed crops with healthy vegetable oils. *Biochimie,* 178, 4-14. https://doi.org/10.1016/j.biochi.2020.09.020

Muñoz-Sanz, J.V., Zuriaga, E., Cruz-García, F., McClure, B. & Romero, C. (2020). Self-(in)compatibility systems: target traits for crop-production, plant breeding, and biotechnology. *Front Plant Sci,* 11, 195. https://doi.org/10.3389/fpls.2020.00195

Murthy, H.N., Hiremath, S.C. & Salimath, S.S. (1993). Origin, evolution and genome differentiation in *Guizotia abyssinica* and its wild species. *Theoretical and Applied Genetics,* 87(5), 587-592. https://doi.org/10.1007/BF00221882

Nadeem, M.A., Nawaz, M.A., Shahid, M.Q., Doğan, Y., Comertpay, G., Yıldız, M., Hatipoğlu, R., Ahmad, F., Alsaleh, A., Labhane, N., Özkan, H., Chung, G. & Baloch, F.S. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment,* 32(2), 261-285. https://doi.org/10.1080/13102818.2017.1400401

Nagella, P., Hosakatte, N., Ravishankar, K., Hahn, E.-J. & Paek, K. (2008). Analysis of genetic diversity among Indian niger [*Guizotia abyssinica* (L. f.) Cass.] cultivars based on randomly amplified polymorphic DNA markers. *Electronic Journal of Biotechnology,* 11(1), 14. https://doi.org/10.2225/vol11-issue1-fulltext-14

Naik, H.K. & Varadahalli, R.D. (2020). Genomic identification of salt induced microRNAs in niger (*Guizotia abyssinica* Cass.). *Plant Gene,* 23, 100242. https://doi.org/https://doi.org/10.1016/j.plgene.2020.100242

Nair, R. & Pandey, M. (2021). Role of molecular markers in crop breeding: a review. *Agricultural Reviews,* 15(6), 171-175. https://doi.org/10.18805/ag.R-2322

Nie, Z.L., Funk, V.A., Meng, Y., Deng, T., Sun, H. & Wen, J. (2016). Recent assembly of the global herbaceous flora: evidence from the paper daisies (Asteraceae: Gnaphalieae). *New Phytol,* 209(4), 1795-806. https://doi.org/10.1111/nph.13740

Nieto, G. & Lorenzo, J.M. (2022). Plant source: Vegetable oils. In: Lorenzo, Munekata, Pateiro, Barba & Domínguez (eds) *Foods Lipids. Sources,*

Health Implications, and Future Trend. Ourense Elsevier. 69-85. https://doi.org/https://doi.org/10.1016/B978-0-12-823371-9.00011-3
Ohri, D. & Khoshoo, T.N. (1987). Nuclear DNA contents in the genusFicus (Moraceae). *Plant Systematics and Evolution,* 156(1), 1-4. https://doi.org/10.1007/BF00937196
Ortiz, R. (2023). Challenges for crop improvement. *Emerg Top Life Sci*, 20230106. https://doi.org/10.1042/etls20230106
Osman, A. (2020). Cold pressed niger (*Guizotia abyssinica* Cass.) seed oil. In: Ramadan (ed.) *Cold Pressed Oils: Green Technology, Bioactive Compounds, Functionality, and Applications*. London: Elsevier. 385-390. https://doi.org/https://doi.org/10.1016/B978-0-12-818188-1.00035-9
Palchamy, K., Kumar, C.H., P S, B., Geethanjali, S., Reddy, Y., Rahul, D. & Senapathy, S. (2022). Current scenario of marker-assisted selection in breeding of minor oilseed crops of India. *J. Oilseeds Res.,* 38, 303-319.
Panda, D., Behera, P.K., Padhi, S.K., Panda, A. & Nayak, J.K. (2023). Neglected and underutilized food plants of India. In: Ismail, Akhtar & Lazarte (eds) *Neglected Plant Foods of South Asia: Exploring and Valorizing Nature to Feed Hunger*. Gewerbestrasse: Springer Nature. 101-118. https://doi.org/10.1007/978-3-031-37077-9_5
Petros, Y., Merker, A. & Zeleke, H. (2007). Analysis of genetic diversity of *Guizotia abyssinica* from Ethiopia using inter simple sequence repeat markers. *Hereditas,* 144(1), 18-24. https://doi.org/https://doi.org/10.1111/j.2007.0018-0661.01969.x
Pham, H., Reisner, J., Swift, A., Olafsson, S. & Vardeman, S. (2022). Crop phenotype prediction using biclustering to explain genotype-by-environment interactions. *Frontiers in Plant Science,* 13, 975976. https://doi.org/10.3389/fpls.2022.975976
Porokhovinova, E.A., Matveeva, T.V., Khafizova, G.V., Bemova, V.D., Doubovskaya, A.G., Kishlyan, N.V., Podolnaya, L.P. & Gavrilova, V.A. (2022). Fatty acid composition of oil crops: genetics and genetic engineering. *Genetic Resources and Crop Evolution,* 69(6), 2029-2045. https://doi.org/10.1007/s10722-022-01391-w
Pour-Aboughadareh, A., Khalili, M., Poczai, P. & Olivoto, T. (2022). Stability indices to deciphering the genotype-by-environment interaction (GEI) effect: an applicable review for use in plant breeding programs. *Plants* 11(3), 414. https://doi.org/10.3390/plants11030414
Prathiba, K. (2020). Analysing the response of non-coding RNA of niger, *Guizotia abyssinica* towards high temperature and associated functional predictions. *Bioscience Biotechnology Research Communications,* 13, 1441-1446. https://doi.org/10.21786/bbrc/13.3/68
Rajaguru, B., Annapurna, C. & Rajeev K, V. (2020). Genetic molecular markers to accelerate genetic gains in crops. *BioTechniques,* 69(3), 158-160. https://doi.org/10.2144/btn-2020-0066

Ramadan, M.F. (2012). Functional properties, nutritional value, and industrial applications of niger oilseeds (*Guizotia abyssinica* Cass.). *Crit Rev Food Sci Nutr,* 52(1), 1-8. https://doi.org/10.1080/10408398.2010.486083

Ranganatha, A.R.G., Panday, A.K., Bisen, R., Jain, S. & Sharma, S. (2016). Niger. In: Gupta (ed.) *Breeding Oilseed Crops for Sustainable Production: Opportunities and Constraints*. London: Elsevier. 169-199. https://doi.org/https://doi.org/10.1016/B978-0-12-801309-0.00008-2

Rebollo, I., Aguilar, I., Pérez de Vida, F., Molina, F., Gutiérrez, L. & Rosas, J.E. (2023). Genotype by environment interaction characterization and its modeling with random regression to climatic variables in two rice breeding populations. *Crop Science,* 63(4), 2220-2240. https://doi.org/https://doi.org/10.1002/csc2.21029

Riley, K.W. & Belayneh, H. (1989). Niger seed : *Guizotia abyssinica* Cass. In: Gerhard, Downey & Ashri, A. (eds) *Oil Crops of the World: Their Breeding and Utilization*. New York: McGraw-Hill. 394-403.

Sadiyah, H. & Hadi, A. (2023). Outlying Observation in Stability Analysis of Genotype: AMMI vs Huehn Method. In: *The 5th International Conference on Agriculture and Life Science 2021 (Icals 2021): "Accelerating Transformation in Industrial Agriculture through Sciences Implementation"*, January 2023, University of Jember. AIP Publishing. 9. https://doi.org/10.1063/5.0117661

Sadras, V.O., Rebetzke, G.J. & Edmeades, G.O. (2013). The phenotype and the components of phenotypic variance of crop traits. *Field Crops Research,* 154, 255-259. https://doi.org/https://doi.org/10.1016/j.fcr.2013.10.001

Sagun, J.V., Yadav, U.P. & Alonso, A.P. (2023). Progress in understanding and improving oil content and quality in seeds. *Frontiers in Plant Science,* 14. https://doi.org/10.3389/fpls.2023.1116894

Samarth, N. & Mahanwar, P. (2015). Modified Vegetable Oil Based Additives as a Future Polymeric Material—Review. *Open Journal of Organic Polymer Materials,* 05, 1-22. https://doi.org/10.4236/ojopm.2015.51001

Sari, P. (2006). *Preliminary design and construction of a prototype canola seed oil extraction machine.(Master thesis 2006).* Ankara, Turkey: Middle East Technical University.

Savva, S. & Kafatos, A. (2016). Vegetable oils: dietary importance. In: *Encyclopedia of Food and Health.* Cyprus: Elsevier Ltd. 365-372.

Schmutterer, H. (2009). Contribution to the knowledge of the crop pest fauna in Ethiopia. *Journal of Applied Entomology,* 67(4), 371-389. https://doi.org/http://dx.doi.org/10.1111/j.1439-0418.1971.tb02136.x

Seegeler, C.J.P. (1983). *Oil plants in Ethiopia, their taxonomy and agricultural significance*. (4). Wageningen: Centre for Agricultural Publishing and Documentation.

Seethapathy, P., Gurudevan, T. & Prabakar, K. (2023). Diseases of Niger (*Guizotia abyssinica* Cass) and their integrated management. In: Bahadur & Dutta

(eds) *Diseases of Oil Crops and Their Integrated Management*. London: CRC Press. 295-301. https://doi.org/10.1201/9781003084198-33

Shahbazi, E. (2019). Genotype selection and stability analysis for seed yield of *Nigella sativa* using parametric and non-parametric statistics. *Scientia Horticulturae,* 253, 172-179. https://doi.org/https://doi.org/10.1016/j.scienta.2019.04.047

Sharma, M., Gupta, S.K. & Mondal, A.K. (2012). Production and trade of major world oil crops. In: Gupta (ed.) *Technological Innovations in Major World Oil Crops*. (1). New York: Springer. 1-15. https://doi.org/10.1007/978-1-4614-0356-2_1

Sharma, S. & Goyal, R. (2015). Effect of environment on fatty acid composition and tocopherol content in soybean, *Glycine max* L Merrill. *Proceedings of the National Academy of Sciences, India Section B: Biological Sciences,* 85(2), 443-450. https://doi.org/10.1007/s40011-014-0354-3

Sharma, S.K., Barthwal, R., Saini, D. & Rawat, N. (2022). Chemistry of food fats, oils, and other lipids. In: Chauhan (ed.) *Advances in Food Chemistry: Food Components, Processing and Preservation*. Mysore: Sprinegr Nature. 209-254. https://doi.org/10.1007/978-981-19-4796-4_6

Shojaei, S.H., Mostafavi, K., Omrani, A., Omrani, S., Nasir Mousavi, S.M., Illés, Á., Bojtor, C. & Nagy, J. (2021). Yield stability analysis of maize (*Zea mays* L.) hybrids using parametric and AMMI methods. *Scientifica,* 2021, 5576691. https://doi.org/10.1155/2021/5576691

Singh, M., Nara, U., Kumar, A., Thapa, S., Jaswal, C. & Singh, H. (2022). Enhancing genetic gains through marker-assisted recurrent selection: from phenotyping to genotyping. *Cereal Research Communications,* 50(3), 523-538. https://doi.org/10.1007/s42976-021-00207-4

Subedi, U., Jayawardhane, K.N., Pan, X., Ozga, J., Chen, G., Foroud, N.A. & Singer, S.D. (2020). The potential of genome editing for improving seed oil content and fatty acid composition in oilseed crops. *Lipids,* 55(5), 495-512. https://doi.org/10.1002/lipd.12249

Sujatha, M., Suresh, G., Helan, T.B. & Pushpa, H.D. (2023). Association of *Golovinomyces ambrosiae* and *Podosphaera xanthii* in causing powdery mildew disease of *Guizotia abyssinica* L.f. Cass in India. *Australasian Plant Pathology,* 52(2), 107-119. https://doi.org/10.1007/s13313-022-00901-9

Tadele, Z. (2018). African orphan crops under abiotic stresses: challenges and opportunities. *Scientifica* 2018, 1451894. https://doi.org/10.1155/2018/1451894

Tanin, M.J., Sharma, A., Saini, D.K., Singh, S., Kashyap, L., Srivastava, P., Mavi, G.S., Kaur, S., Kumar, V., Kumar, V., Grover, G., Chhuneja, P. & Sohu, V.S. (2022). Ascertaining yield and grain protein content stability in wheat genotypes having the Gpc-B1 gene using univariate, multivariate, and correlation analysis. *Frontiers in Genetics,* 13, 1001904. https://doi.org/10.3389/fgene.2022.1001904

Terefe, G., Gorfu, D., Degefu, D. & Abebe, F. (2008). Review of research on diseases of oil crops in Ethiopia. In: Abraham (ed.) *Increasing Crop Production Through Improved Plant Protection*. Addis Abeba: PPSE. 253-273. https://www.researchgate.net/publication/261015723

Terefe, M., Birmeta, G., Girma, D., Geleta, M. & Tesfaye, K. (2023). Analysis of genetic diversity and population structure of oilseed crop noug (*Guizotia abyssinica*) accessions collected from Ethiopia. *Molecular Biology Reports,* 50(1), 43-55. https://doi.org/10.1007/s11033-022-08005-9

Terefe, M. & Girma, D. (2022). Development of molecular resources for the genetic improvement of noug (*Guizotia abyssinica* (L.f) Cass): a mini review. *CABI Agriculture and Bioscience,* 3(1), 52. https://doi.org/10.1186/s43170-022-00121-7

Tesfaye, M. (2019). Understanding the phenotypic diversity and ecological distribution of noug (*Guizotia abyssinica*) for its improvement and conservation. *Journal of Natural Sciences Research,* 9(5), 38-46.

Thatte, P. & Jyothi Lakshmi, A. (2012). Nutritional potential, bioaccessibility of minerals and antioxidant properties of niger (*Guizotia abyssinicacass*) seed protein. *International Journal of Food Science & Technology,* 47(3), 656-663. https://doi.org/https://doi.org/10.1111/j.1365-2621.2011.02890.x

Thudi, M., Palakurthi, R., Schnable, J.C., Chitikineni, A., Dreisigacker, S., Mace, E., Srivastava, R.K., Satyavathi, C.T., Odeny, D., Tiwari, V.K., Lam, H.-M., Hong, Y.B., Singh, V.K., Li, G., Xu, Y., Chen, X., Kaila, S., Nguyen, H., Sivasankar, S., Jackson, S.A., Close, T.J., Shubo, W. & Varshney, R.K. (2021). Genomic resources in plant breeding for sustainable agriculture. *Journal of Plant Physiology,* 257, 153351. https://doi.org/https://doi.org/10.1016/j.jplph.2020.153351

Tomé-Rodríguez, S., Barba-Palomeque, F., Ledesma-Escobar, C.A., Miho, H., Díez, C.M. & Priego-Capote, F. (2023). Influence of genetic and interannual factors on the fatty acids profile of virgin olive oil. *Food Chemistry,* 422, 136175. https://doi.org/https://doi.org/10.1016/j.foodchem.2023.136175

Tsehay, S., Ortiz, R., Geleta, M., Bekele, E., Tesfaye, K. & Johansson, E. (2021). Nutritional profile of the Ethiopian oilseed crop noug (*Guizotia abyssinica* Cass.): opportunities for its improvement as a source for human nutrition. *Foods,* 10(8), 1778. https://doi.org/10.3390/foods10081778

Tsehay, S., Ortiz, R., Johansson, E., Bekele, E., Tesfaye, K., Hammenhag, C. & Geleta, M. (2020). New transcriptome-based SNP markers for noug (*Guizotia abyssinica*) and their conversion to KASP markers for population genetics analyses. *Genes* 11(11), 1373. https://doi.org/10.3390/genes11111373

Ur Rehman, S., Ali Sher, M., Saddique, M.A.B., Ali, Z., Khan, M.A., Mao, X., Irshad, A., Sajjad, M., Ikram, R.M., Naeem, M. & Jing, R. (2021). Development and Exploitation of KASP Assays for Genes Underpinning

Drought Tolerance Among Wheat Cultivars From Pakistan. *Frontiers in Genetics,* 12, 684702. https://doi.org/10.3389/fgene.2021.684702

Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q. & Sorrells, M.E. (2021). Designing future crops: genomics-assisted breeding comes of age. *Trends Plant Sci,* 26(6), 631-649. https://doi.org/10.1016/j.tplants.2021.03.010

Vaughan, J.G. (1970). *The structure and utilization of oil seeds*. London Chapman & Hall Ltd. https://search.library.wisc.edu/catalog/999627005102121

Waseem, S., Imadi, S., Gul, A. & Ahmad, P. (2017). *Oilseed crops: yield and adaptations under environmental stress*. Chichester,: John Wiley & Sons Ltd. https://doi.org/10.1002/9781119048800.ch1

Weiss, E.A. (1983). *Oilseed Crops*. (Tropical Agriculture). London Longman. https://doi.org/10.1017/S0014479700017865

Whitley, E. & Ball, J. (2002). Statistics review: nonparametric methods. *Crit Care,* 6(6), 509-13. https://doi.org/10.1186/cc1820

Wolf, R.B., Cavins, J.F., Kleiman, R. & Black, L.T. (1982). Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. *Journal of the American Oil Chemists' Society,* 59(5), 230-232. https://doi.org/10.1007/BF02582182

Yadav, S., Kumar, S., Hussain, Z., Suneja, P., Yadav, S.K., Nizar, M.A. & Dutta, M. (2012). *Guizotia abyssinica* (L.f.) cass.: an untapped oilseed resource for the future. *Biomass and Bioenergy,* 43, 72-78. https://doi.org/https://doi.org/10.1016/j.biombioe.2012.03.025

Yousaf, M.A., Akhtar, N., Gul, S., Saleem, U., Mahmood, T., Aziz, A., Asif, M. & Raza, N. (2022). Genotype by environment interaction and association of yield contributing traits in sunflower genotypes under the environmental condition of Sargodha, Pakistan. *Helia,* 45(76), 55-70. https://doi.org/doi:10.1515/helia-2021-0019

Zemour, K., Adda, A., Labdelli, A., Dellal, A., Cerny, M. & Merah, O. (2021). Effects of genotype and climatic conditions on the oil content and its fatty acids composition of *Carthamus tinctorius* L. seeds. *Agronomy,* 11(10), 2048. https://www.mdpi.com/2073-4395/11/10/2048

Zhang, C., Huang, C.-H., Liu, M., Hu, Y., Panero, J.L., Luebert, F., Gao, T. & Ma, H. (2021). Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. *Journal of Integrative Plant Biology,* 63(7), 1273-1293. https://doi.org/https://doi.org/10.1111/jipb.13078

Zhou, Y., Zhao, W., Lai, Y., Zhang, B. & Zhang, D. (2020). Edible plant oil: global status, health issues, and perspectives. *Frontiers in Plant Science,* 11, 1315. https://doi.org/10.3389/fpls.2020.01315

Popular science summary

Noug, an oilseed crop known scientifically as *Guizotia abyssinica* and niger seed by its common name, is mainly grown in Ethiopia and India for its oil and seeds. Its primary center of origin and diversity is in Ethiopia. It is renowned for its nutritional richness and diverse applications. In Ethiopia, it holds a pivotal role, significantly bolstering the livelihoods of numerous small-scale farmers while playing a crucial part in ensuring food security and income generation. Noug seeds are a treasure trove of essential nutrients, boasting proteins, healthy fats, vitamins, and minerals. The oil extracted from these seeds is not only a staple in Ethiopian kitchens, providing sustenance and energy for the population but also has a reputation for potential health benefits. Traditionally, it has found its way into Ethiopian folk medicine, believed to possess medicinal properties such as anti-inflammatory and antioxidant effects. Furthermore, as a valuable export commodity, noug contributes significantly to the country's foreign exchange earnings, solidifying its status as a vital agricultural asset.

The indigenous genetic resource of noug in Ethiopia is highly genetically diverse therefore, it has a high potential for improvement by breeding and biotechnology. Noug is one of the underutilized and research-neglected crops with little scientific research. Little efforts have been made to generate better cultivars with desirable characteristics such as high oil content, oil quality, yield, disease resistance and other traits. In noug, populations there are bottlenecks such as self-incompatibility, shattering, lodging, low yield and less attention for agronomic management, which hinders its improvement. There

is lack of thorough research on genomics, transcriptomics, proteomics, mapping of genes and quantitative trait loci, and for developing self-compatible lines to enhance desirable traits compared to other oilseeds such as sunflower and brassicas.

In this research, we included diverse genetic resources for nutritional profiling of lipids, fatty acids, protein and minerals, genetic diversity, population genetics analysis, genomic tools development and the effect of genotype and environment on some desirable traits such as oil content and major fatty acids such as linoleic, oleic, palmitic and stearic acids quality traits. In this study, we found that the content of noug seed oil ranged from 33 to 46% and fatty acids linoleic (72–79%), oleic (5–9%), palmitic (8–10%) and stearic acid (7–10%). The seed protein content was between 25 and 28%. Noug is also a good source of minerals though it is affected by the location where the genotype grew. We found potassium (K) being the highest 9254 (μg g-1) among the studied macro and micro mineral nutrient elements. The development of genomic tools is essential for understanding genetic variability and for further facilitating breeding and conservation. In this research a noble transcriptome-based single nucleotide polymorphism marker known as PCR-based KASP was developed and utilized for population genetics analyses. The development of these genomic tools was useful for population genetic diversity analysis. According to these markers, near to 50 % of loci revealed significant deviation from Hardy-Weinberg equilibrium when they were validated using noug accession. The markers also showed a high genetic variation within accessions ranging from 90–95%. The developed genomic tool will have applications in future noug research. Genotype-by-environment interaction research is common nowadays. Having cultivars that are tailored to a specific environment or genotype that performs better in all environments are breeding challenge. We have evaluated the effect of genotype by environment interaction (GEI) on oil content and major fatty acids of noug. The stability of each genotype for oil content and major fatty acids was analyzed using the additive main effect and multiplicative interactions (AMMI) model, and both Eberhart-Russel and Finlay-Wilkinson regression methods. The AMMI model revealed that 25%, 57%, and 49% of the total variation in oil content, oleic acid and linoleic acids, respectively, were due to GEI. The first two principal component interaction axes (IPCA) for the GEI accounted

for 88%, 72% and 78% of the total GEI variance for oil content, oleic and linoleic acids, respectively. Hence, AMMI models were able to determine the effect of GEI on the variation in oil content and major fatty acids in noug. Stability analysis revealed that some noug genotypes were less affected by the testing environments. Eberhart-Russel and Finlay-Wilkinson regression models were also able to identify the most stable noug genotypes for oil quantity and quality. This research provides useful knowledge for the selection and breeding of superior noug cultivars in Ethiopia.

In brief, the findings of this PhD thesis research have proven to have practical relevance in the domains of pre-breeding assessment, breeding practices, and the preservation of Ethiopian noug genetic resources.

Populärvetenskaplig sammanfattning

Noug, en oljeväxt som är vetenskapligt känd som *Guizotia abyssinica* och vid sitt vanliga namn nigerfrö, odlas främst i Etiopien och Indien för dess olja och frön. Dess huvudsakliga ursprung och center för mångfald är i Etiopien. Den är känd för sitt goda näringsinnehåll och har en rad olika användningsområden. I Etiopien spelar den en avgörande roll genom att stärka småskaliga bönders uppehälle samtidigt som den är viktig för landets livsmedelssäkerhet och inkomstgenerering. Nougfrön är en skattkammare av livsviktiga näringsämnen och innehåller proteiner, nyttiga fetter, vitaminer och mineraler. Oljan som utvinns från dessa frön är inte bara en stapelvara i etiopiska kök som ger näring och energi till befolkningen, utan har också ett rykte om potentiella hälsofördelar. Traditionellt har den hittat sin plats inom etiopisk folkmedicin och anses ha medicinska egenskaper såsom antiinflammatoriska och antioxidanta effekter. Som en värdefull exportvara bidrar noug dessutom i hög grad till landets utländska valuta, vilket befäster dess status som en viktig resurs för det etiopiska jordbruket.

De inhemska genetiska resurserna av noug i Etiopien är av hög genetisk diversitet och har därför stor potential i att bidra till en förbättring av noug genom växtförädling och bioteknik. Noug är en av många underutnyttjade och försummade grödor där mycket begränsad vetenskaplig forskning hittills har bedrivits. Begränsade ansträngningar har gjorts för att generera bättre sorter med önskvärda egenskaper som högt oljeinnehåll, god oljekvalitet, hög avkastning, resistens mot sjukdomar samt andra egenskaper. Inom nougpopulationer finns egenskaper som behöver förbättras, såsom självinkompatibilitet, spridning av frön, stråläggning, låg avkastning och bristande agronomiska brukarmetoder, vilket hindrar dess produktivitet. Det saknas också

omfattande forskning om genomik, transkriptomik, proteomik, kartläggning av gener och kunskap om egenskapers genetiska kontroll, samt utveckling av självkompatibla linjer för att förbättra önskvärda egenskaper.

I denna forskning inkluderades olika genetiska resurser för näringsprofilering av lipider, fettsyror, protein och mineraler, samt studier av genetisk mångfald och populationsgenetik. Vi utvecklade även genomiska verktyg och- studerade effekten av genotyp och miljö för några önskvärda egenskaper såsom oljeinnehåll och viktiga fettsyror som linolsyra, oljesyra, palmitinsyra och stearinsyra. I denna studie fann vi att innehållet av nougfröolja varierade från 33 till 46% och fettsyrorna linolsyra (72–79%), oljesyra (5–9%), palmitinsyra (8–10%) och stearinsyra (7–10%). Proteininnehållet i fröna låg mellan 25 och 28%. Noug är också en god källa för mineraler, även om det påverkas av platsen där plantorna växte. Vi fann att kalium (K) var det högsta med 9254 ($\mu g\ g^{-1}$) bland de studerade makro- och mikromineralämnen. Utvecklingen av genomiska verktyg är avgörande för att förstå genetisk variation och för att möjliggöra vidare förädling och bevarande. I denna forskning utvecklades även nya SNP-markörer via transkriptomdata för PCR-baserad KASP analys som användes för populationsgenetiska analyser. Dessa genomiska verktyg visade sig vara användbara för analys av genetisk mångfald inom den studerade nougpopulationen. Enligt denna analys visade nästan 50 % av loci signifikant avvikelse från Hardy-Weinbergjämvikt när de validerades med nougaccessioner. Markörerna påvisade också en hög genetisk variation inom accessioner, som sträckte sig från 90 till 95%. De utvecklade genomiska verktygen kommer att ha tillämpningar inom framtida nougforskning. Genotyp-miljö-interaktion är en vanlig analys inom forskningsfältet. Att ha sorter som är anpassade till en specifik miljö eller genotyper som presterar bättre i alla miljöer är en utmaning inom växtförädlingen. Vi utvärderade effekten av genotyp-miljö-interaktion (GEI) på oljeinnehåll och viktiga fettsyror hos noug. Stabiliteten för varje genotyp för oljeinnehåll och viktiga fettsyror analyserades med hjälp av additiv huvudeffekt och multiplikativa interaktioner (AMMI)-modellen samt Eberhart-Russel och Finlay-Wilkinson regressionsmetoder. AMMI-modellen visade att 25 %, 57 % och 49 % av den totala variationen i oljeinnehåll, oljesyra och linolsyror, respektive, berodde på GEI. De första två huvudkomponenternas interaktionsaxlar (IPCA) för GEI svarade för 88 %, 72 % och 78 % av den totala GEI-variationen för oljeinnehåll, oljesyra och linolsyror, respektive.

Därför kunde vi med hjälp av AMMI-modellerna bestämma effekten av GEI på variationen i oljeinnehåll och viktiga fettsyror hos noug. . Stabilitetsanalyser avslöjade att vissa nouggenotyper påverkades mindre av miljöeffekter. Med Eberhart-Russel och Finlay-Wilkinson regressionsmodeller kunde vi också identifiera de mest stabila nouggenotyperna avseende oljemängd och kvalitet. Denna forskning bidrar med användbar kunskap för urval och förädling av förbättrade nougsorter i Etiopien.

Sammanfattningsvis har resultaten av denna doktorsavhandling visat sig ha relevans inom områdena för-förädling (pre-breeding), praktisk förädling och bevarande av genetiska resurser av etiopisk noug.

Acknowledgements

I put my trust in the almighty God! God, son of the Virgin Mary, receives the lion's share of my thanks. I would like to express my heartfelt thanks to all institutions, research funding bodies/organizations and individuals who contributed to the success of my PhD thesis completion. During my PhD study, I had the chance to know and interact with many people who have offered me support and important advice. Thanks to all as a team and individual help towards the end, which justifies the means.

Supervisory team composition was vibrant and included Prof. Rodomiro Ortiz, Assoc. Prof. Mulatu Geleta and Prof. Eva Johansson from the Department of Plant Breeding (VF) Swedish University of Agricultural Sciences (SLU) as well as Prof. Endashaw Bekele and Assoc. Prof. Kassahun Tesfaye from Addis Ababa University. My immediate thanks go to my main supervisor, Prof. Rodomiro Ortiz, for your unreserved guidance, suggestions, and supervision from the start of my study until now. You are always open to me and easy to discuss at any time with no prior appointment. The academic freedom you cherish, while you are keeping academic professionalism and quality, is memorable in a lifetime. I never forget your immediate action to buy a new laptop when I lost mine on public transport in Malmö, Sweden. Assoc. Prof. Mulatu Geleta: I am astonished by your hard work and commitment. Thank you for the guidance through each stage of my study journey from fund acquisition to completion. Field, greenhouse, laboratory work, academic encouragement and constructive suggestions during the PhD process

are few to mention. Prof. Eva Johansson: I would like to extend my sincere thanks to you. You had a great role in supervising my PhD. Your professional efforts especially in the edition, revision and correction of scientific papers are unforgettable. Prof. Endashaw Bekele: You are one of my esteemed supervisors for your unwavering support and belief in me. When I have difficulties in fieldwork in Ethiopia, you are the one to solve the issues related to technical and financial support. Your critical paper review is one that I must mention. Assoc. Professor Kassahun Tesfaye: your immense knowledge and plentiful experience have encouraged me all the time in my academic research career. Historically, you were my supervisor during my master's study. You have been a sincere scientific advisor since my first contact. Your insights on manuscript revisions were deep. I am also deeply grateful to Prof. Anders Carlsson for your genuine support from the get-go of my PhD study admission to completion. To be frank, now I feel as if you are one of my academic supervisors instead of the administration part. Prof. Gurja Belay, Assoc. Prof. Adey Feleke and Asst. Prof. Helen Nigussie who are my colleagues in the Microbial, Cellular and Molecular Biology Department at AAU, played a great role in facilitating administrative issues such as study leave and extension. Heartfelt thanks for this.

Hello! I would like to mention two truly exceptional people from SLU, Alnarp with respect to gas chromatography technical help with no normal call of their duty to my thesis work. These researchers are Asst. Prof. Åsa Grimberg and Assoc. Prof. Ida Lager. You both equally deserve acknowledgment for the free lunch technical support. Assoc. Prof. Cecilia Hammenhag, your help in translating the abstract and popular science summary of my PhD thesis from English to Swedish is sincerely appreciate. I would like to take this opportunity to thank Asst. Prof. Jan-Eric Englund and Mr. Adam Flöhr for your excellent assistance in teaching me how to interpret statistical data.

My gratitude extends to the Swedish Research Council (Vetenskapsrådet, VR) for the funding opportunity to undertake my PhD study at the Department of Plant Breeding, SLU, Alnarp. I am also deeply indebted to my home institution, Addis Ababa University, for granting me study leave and the opportunity to join SLU in Sweden as PhD student. Holeta Agricultural Re-

search Center in Ethiopia is highly acknowledged for providing field experimental sites and providing technical and labour works. Mr. Misteru Tesfaye played a leading role, fingers crossed.

In achieving the successful completion of my doctoral study, multitudes of individuals in diverse roles have contributed. I extend my gratitude to those affiliated with the following institutions: Addis Ababa University (AAU), Holeta Agricultural Research Center (HARC), and Swedish University of Agricultural Sciences (SLU).

AAU: Asst. Prof. Addisu Mekonnen, Assoc. Prof. Asnake Desalegn, Asst. Prof. Bayable Atmnafu, Assoc. Prof. Dereje Beyene, Assoc. Prof. Fitsum Tigu, Asst. Prof. Mekibib Fekadu, Asst. Prof. Tesfaye Admasu, Kidist Cherkose, late Assoc. prof. Kifle Dagne, Shewangizaw Abay, Temesgen Dagnaw and Prof. Tileye Feyissa.

HARC: Balcha regassa, Birhanu Mengistu, Samra Daniel, Sinafikish H/Michael and Tadesse Debele.

SLU: Birhanu Worabo, Camilla Stjarnang, Desirée Lindqvist, Dessie Yirdaw, Asst. Prof. Abel Teshome, Asst. Prof. Adane Gebeyehu, Asst. Prof. Admas Alemu, Asst. Prof. Betelehem Wondwosen, Prof. Tomas Bryngelsson, Asst. Prof. Joel Markgren, Asst. Prof. Mahbubjon Rahmatov, Asst. Prof. Tibebe Dejene, Helen Lindgren, Malin Olsson, Mirela Beganovic, Rimsha Ashraf, Wubbadis Bekele and Zerihun Senbeto. The working condition was vibrant in the Department of Plant Breeding at SLU. I thank all staff and friends of SLU for their warm and welcoming faces.

Finally, yet importantly, I deeply thank my sweet wife Rekik Tizazu and adorable sons Matias, Natanim and Henok Sewalem who are a constant inspiration to me. Thank you for all and each one of the wonderful moments that you gave me here in Sweden and unpleasant distance. Your inspiration remains my main support throughout my career and endeavors.

I

foods

MDPI

Article

Nutritional Profile of the Ethiopian Oilseed Crop Noug (*Guizotia Abyssinica* Cass.): Opportunities for Its Improvement as a Source for Human Nutrition

Sewalem Tsehay [1], Rodomiro Ortiz [1], Mulatu Geleta [1,*], Endashaw Bekele [2], Kassahun Tesfaye [2] and Eva Johansson [1]

[1] Department of Plant Breeding, Swedish University of Agricultural Sciences, P.O. Box 190, SE-23422 Lomma, Sweden; sewalem.tsehay@slu.se (S.T.); rodomiro.ortiz@slu.se (R.O.); Eva.Johansson@slu.se (E.J.)

[2] Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, Addis Ababa P.O. Box 1176, Ethiopia; endashawbw@gmail.com (E.B.); kassahuntesfaye@yahoo.com (K.T.)

* Correspondence: mulatu.geleta.dida@slu.se

Abstract: The aim of this study was to evaluate the potential of noug as a source for human nutrition. Diverse noug genotypes were evaluated for their content and/or composition of total lipids, fatty acids, proteins, and minerals using standard methods. The total lipid content (32.5–45.7%) and the proportion of an essential fatty acid, linoleic acid (72.2–77.8%), were high in noug, compared to other oilseed crops. The proportion of oleic acid, a monounsaturated fatty acid, was low in noug (5.2–9.2%). The breeding objective of increasing the oleic acid level in the highland, where noug is mainly cultivated, was limited, as the content of this acid was low in this environment. The seed protein concentration (25.4–27.5%) and mineral content were mainly affected by the cultivation environment, as the high temperature increased the amount of protein, whereas the soil condition was a major factor in the variation of the mineral content. Thus, noug is a unique crop with a high seed oil content, of which a high proportion is linoleic acid. With the exception of the seed oleic acid content, when grown in low-altitude areas, the genotypic variation contributes less than the cultivation environment to the nutritional attributes of noug. Hence, high-oleic-acid noug for lowland production can be targeted as a breeding goal.

Keywords: *Guizotia abyssinica*; lipids; mineral elements; noug; oilseeds; protein

Citation: Tsehay, S.; Ortiz, R.; Geleta, M.; Bekele, E.; Tesfaye, K.; Johansson, E. Nutritional Profile of the Ethiopian Oilseed Crop Noug (*Guizotia Abyssinica* Cass.): Opportunities for Its Improvement as a Source for Human Nutrition. *Foods* **2021**, *10*, 1778. https://doi.org/10.3390/foods10081778

Academic Editor: Marcello Iriti

Received: 10 July 2021
Accepted: 29 July 2021
Published: 31 July 2021

1. Introduction

Oilseeds have the potential to contribute a vital source of nutrition to the human diet [1]. Vegetable oils, proteins, and minerals are components present in oilseeds, all having essential functions for the human body [1,2]. The profile of the content and composition of nutrients in a specific oilseed vary based on the species, type, maturity, environmental factors, breeding objectives, and management [3,4].

Currently, oilseeds are primarily used for oil extraction, and they are considered as the main sources of vegetable oil [5]. The quality of the seed oil is mainly determined by its fatty acid composition [6], which is known to have an impact on human health; e.g., soybean, corn, nuts, sunflower, safflower, and noug seed oils are rich in omega-6 fatty acids, which play a role in preventing cardiovascular disorders [7]. The oil extraction also results in a side product in the form of a press-cake, which is mainly used as animal feed, where its nutritional composition plays a role [8,9]. However, traditionally, oilseeds are consumed as a wholegrain food component. For example, in Ethiopia, slightly roasted noug (*Guizotia abyssinica* Cass.) seeds are finely pounded using a mortar and pestle, until a thick fluid, locally known as *litlit*, is formed. Then, bread made from different cereals, sugar, salt, etc., are added to the *litlit* and further pounded. The mixture is then made into balls of different sizes and served as a food locally known as *chifko* [10]. The current growing global

population calls for an increased use of whole-seed oilseeds and/or oilseed cakes as a human nutrient source due to their high nutritive profile [11,12]. The proteins from oilseeds are known to have a high content of essential amino acids and are therefore beneficial to human health and well-being [13]. The oilseed minerals [14] have the potential to play an essential role in the human body by providing both macro- and micronutrients [15]. Mineral intake in sufficient amounts is essential for a vital life in humans and animals, and the consumption of inadequate amounts may result in an inefficient structure of muscles and the malfunction of nerves and metabolic processes, contributing to, e.g., threatened immunity, cognitive memory, and regulatory functions [16,17].

Noug is one of the major sources of edible seed oil in Ethiopia [18,19]. The vegetable seed oil from noug contributes significant economic and nutritional values to the country and its population, and it is ranked as a superior seed oil among Ethiopians [20]. Noug seeds have a relatively high trade value, contributing income to Ethiopia through its export [21]. A total lipid content of 25–56% has been shown for noug grown in Ethiopia [22–27]. The main fatty acids in noug oil are linoleic acid (C18:2), oleic acid (C18:1), palmitic acid (C16:0), and stearic acid (C18:0), contributing more than 90% of the total fatty acids in the seed oil [27]. The protein content of whole-seed noug has reported been limited, although seed meal of Ethiopian noug has been reported to contain about 30% protein [28]. A proximate analysis of noug seeds grown in the USA showed a higher level of protein content (28.2%) than imported noug seeds from Ethiopia (18.3%) and India (26.6%) [19]. A few studies have also evaluated the content of macro- and microminerals in the noug seed, and variations of these elements have been reported [29–31]. While several studies have been conducted on noug seed oil content and fatty acid composition, there is a lack of broad investigations covering the full nutritional potential of noug.

Hence, the objective of this paper was to investigate the quantity and quality of the nutrients of noug seeds, including lipids, fatty acids, proteins, and minerals, sampled from a broad range of materials, including cultivars, landraces collected from wide geographic areas in Ethiopia, and "breeding populations", selected based on lipid content and yield parameters. The nutritive value of noug and opportunities for production of high-nutritive-value noug in Ethiopia were evaluated, and comparisons of the lipid and fatty acid composition with other major oilseed crops were carried out.

2. Materials and Methods

2.1. Plant Material and Field Experiments

Basically, three different noug materials were used in the present study: (i) twenty-eight landrace populations, (ii) ten "breeding populations", and (iii) seven released cultivars (Table 1). The landrace populations were collected directly from farmers' fields in four Federal States of Ethiopia (Amhara, Oromia, SNNPs and Tigray) during 2003. The altitude of origin of these samples (in meters above sea level) are given in Table 1. The seeds were rejuvenated in 2008 and 2012 by growing them in a greenhouse. The 10 breeding populations were developed through four generations of crossbreeding, selecting them based on their lipid content, oleic acid content, and/or seed size [27]. The seeds of the seven cultivars were obtained from Holeta Agricultural Research Center (HARC) of the Ethiopian Institute of Agricultural Research (EIAR). For the sake of simplicity, each population and cultivar will be referred to as a "genotype" from here on. In order to diminish the environmental effects due to differences in previous cultivations and to understand the interactions between genotypic and environmental variation, all the three materials of noug were grown in Ethiopia at three sites: Holeta (09°04′ N, 38°29′ E; 2400 masl), Ginchi (9°01′ N and 38°10′ E; 2300 masl), and Debrezeit (8°44′ N and 38°58′ E; 1900 masl) from June to December 2016 (Figure 1).

Table 1. Least square means (%), obtained from the mixed model ± standard error, of the total oil content and content of the fatty acids of noug genotypes and the altitude of origin of the various samples.

Genotype	Total Lipid Content	Linoleic Acid	Oleic Acid	Palmitic Acid	Stearic Acid	Altitude of Origin (Masl)
NG-83 [b]	39.8 ± 1.6	75.4 ± 1.0	6.4 ± 0.6	9.5 ± 0.3	7.7 ± 0.5	na
NG-84 [b]	41.2 ± 1.6	73.2 ± 1.0	8.3 ± 0.6	8.6 ± 0.3	8.1 ± 0.5	na
NG-85 [a]	32.5 ± 2.7	76.9 ± 1.0	6.7 ± 1.1	8.6 ± 0.5	7.4 ± 0.9	1740
NG-86 [a]	38.1 ± 1.6	74.0 ± 1.0	6.6 ± 0.6	9.5 ± 0.3	9.5 ± 0.5	1830
NG-87 [a]	36.0 ± 1.6	75.4 ± 1.0	5.2 ± 0.6	9.6 ± 0.3	9.3 ± 0.5	1540
NG-88 [a]	35.5 ± 2.7	77.2 ± 1.7	6.2 ± 1.1	8.5 ± 0.5	7.6 ± 0.9	1680
NG-89 [a]	38.1 ± 1.6	75.3 ± 1.0	7.1 ± 0.6	9.0 ± 0.3	8.8 ± 0.5	1890
NG-90 [a]	38.7 ± 1.6	75.7 ± 1.0	6.9 ± 0.6	8.5 ± 0.3	8.5 ± 0.5	2440
NG-91 [a]	38.5 ± 1.9	76.0 ± 1.2	7.4 ± 0.8	8.4 ± 0.4	7.6 ± 0.6	1890
NG-92 [b]	44.4 ± 1.9	76.5 ± 1.2	6.9 ± 0.8	8.6 ± 0.4	7.6 ± 0.6	1860
NG-93 [a]	36.4 ± 1.9	77.8 ± 1.2	5.5 ± 0.8	8.0 ± 0.4	8.1 ± 0.6	2425
NG-94 [a]	34.5 ± 1.9	77.2 ± 1.2	5.9 ± 0.8	8.6 ± 0.4	8.3 ± 0.6	1920
NG-95 [a]	39.0 ± 1.6	74.8 ± 1.0	7.6 ± 0.6	8.3 ± 0.3	9.0 ± 0.5	1968
NG-96 [a]	36.8 ± 2.7	77.3 ± 1.7	6.2 ± 1.1	8.6 ± 0.5	7.3 ± 0.9	1820
NG-97 [a]	39.3 ± 1.6	77.2 ± 1.0	6.6 ± 0.6	8.7 ± 0.3	7.4 ± 0.5	1640
NG-98 [a]	39.2 ± 1.6	75.4 ± 1.0	7.0 ± 0.6	8.4 ± 0.3	8.5 ± 0.5	1590
NG-99 [b]	45.7 ± 1.9	76.4 ± 1.0	7.1 ± 0.8	8.3 ± 0.4	8.5 ± 0.6	2460
NG-101 [b]	40.0 ± 1.9	75.4 ± 1.2	6.8 ± 0.8	8.9 ± 0.4	8.5 ± 0.6	2400
NG-102 [a]	37.0 ± 1.9	74.7 ± 1.2	7.7 ± 0.8	8.7 ± 0.4	8.2 ± 0.6	2430
NG-103 [a]	35.4 ± 1.9	75.2 ± 1.2	7.0 ± 0.8	8.5 ± 0.4	9.2 ± 0.6	2045
NG-105 [a]	34.9 ± 1.6	76.4 ± 0.2	6.4 ± 0.6	8.3 ± 0.3	8.3 ± 0.5	1400
NG-106 [a]	37.8 ± 1.6	76.3 ± 1.0	7.3 ± 0.6	8.5 ± 0.3	8.3 ± 0.5	1840
NG-107 [b]	40.7 ± 1.9	76.3 ± 1.2	6.5 ± 0.8	7.9 ± 0.4	8.8 ± 0.6	1945
NG-108 [a]	33.6 ± 1.6	75.5 ± 1.0	6.9 ± 0.6	9.0 ± 0.3	8.3 ± 0.5	1440
NG-109 [a]	33.1 ± 1.6	74.9 ± 1.0	6.8 ± 0.6	9.0 ± 0.3	8.8 ± 0.5	1400
NG-111 [b]	39.4 ± 1.9	75.8 ± 1.2	7.4 ± 0.8	8.8 ± 0.4	7.2 ± 0.6	2590
NG-112 [a]	35.7 ± 1.9	76.4 ± 1.2	6.4 ± 0.8	8.3 ± 0.4	8.2 ± 0.6	2155
NG-113 [a]	37.2 ± 1.6	72.3 ± 1.0	9.2 ± 0.6	7.8 ± 0.3	10.0 ± 0.5	2210
NG-114 [b]	35.6 ± 1.6	75.2 ± 1.0	7.1 ± 0.6	9.0 ± 0.3	8.5 ± 0.5	na
NG-115 [a]	41.4 ± 1.6	77.2 ± 1.0	7.1 ± 0.6	8.4 ± 0.3	7.8 ± 0.5	1700
NG-117 [b]	43.7 ± 1.9	76.6 ± 1.2	6.1 ± 0.8	8.8 ± 0.4	8.3 ± 0.6	2210
NG-118 [a]	39.2 ± 1.9	76.2 ± 1.2	6.6 ± 0.8	8.5 ± 0.4	8.1 ± 0.6	2540
NG-119 [a]	35.5 ± 2.7	74.5 ± 1.2	7.5 ± 1.1	8.8 ± 0.5	9.3 ± 0.9	2370
NG-120 [a]	39.2 ± 1.6	75.5 ± 1.0	6.7 ± 0.6	8.6 ± 0.3	8.9 ± 0.5	2005
NG-121 [a]	37.8 ± 1.6	75.3 ± 1.0	7.0 ± 0.6	8.4 ± 0.3	8.8 ± 0.5	1650
NG-122 [a]	32.8 ± 2.7	75.9 ± 1.7	6.3 ± 1.1	8.8 ± 0.5	8.9 ± 0.9	1790
NG-123 [a]	38.1 ± 1.6	75.7 ± 1.0	6.9 ± 0.6	8.5 ± 0.3	8.5 ± 0.5	1600
NG-124 [b]	39.1 ± 1.6	75.9 ± 1.0	6.4 ± 0.6	8.7 ± 0.3	8.1 ± 0.5	na
EST [c]	34.5 ± 2.7	73.5 ± 1.7	8.2 ± 1.1	8.6 ± 0.5	9.8 ± 0.9	na
EVE [c]	35.9 ± 1.6	75.2 ± 1.0	7.5 ± 0.6	8.7 ± 0.3	8.6 ± 0.5	na
FOG [c]	34.3 ± 1.6	77.4 ± 1.0	5.8 ± 0.6	8.5 ± 0.3	7.7 ± 0.5	na
GIN [c]	37.4 ± 1.6	75.4 ± 1.0	6.3 ± 0.6	9.0 ± 0.3	8.8 ± 0.5	na
KUY [c]	37.6 ± 1.6	76.5 ± 1.0	6.6 ± 0.6	8.6 ± 0.3	8.0 ± 0.5	na
LVE [c]	33.0 ± 1.6	76.0 ± 1.0	6.2 ± 0.6	8.4 ± 0.3	8.9 ± 0.5	na
SHA [c]	34.6 ± 1.6	76.3 ± 1.0	6.2 ± 0.6	8.5 ± 0.3	8.3 ± 0.5	na

[a] landrace population, [b] breeding population, [c] released cultivar. na = not applicable.

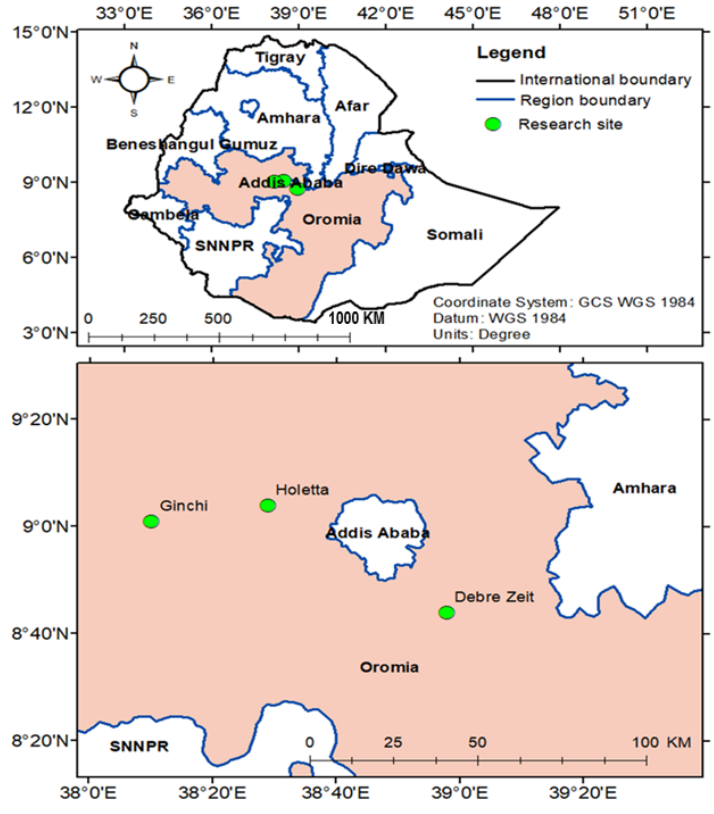


Figure 1. A map showing the locations of the three field experimental sites in Ethiopia.

An incomplete block design within each site was used. Each genotype was planted on a 3 m by 2.1 m (6.3 sqm) plot in seven rows, with a 30 cm distance between rows. A distance of 1 m between the plots was used. A pollination net of 3.2 m (l) × 2.3 m (w) × 2 m (h) was used to cover the plants on each plot, just before flowering, to prevent gene flow between genotypes from different plots. Hand pollination was carried out between plants within each plot by gently rubbing mature flower heads (capitula) in the morning every other day for two weeks. Each genotype and plot was separately harvested at maturity and threshed, and clean seeds were used for laboratory analyses. The lipid content and fatty acid composition were evaluated separately for the seeds from each genotype, plot, and site. Due to resource limitations, 21 of the 28 landrace populations, nine of the 10 "breeding populations", and six of the seven cultivars were used for protein and mineral content analyses.

2.2. Materials and Chemicals Used for the Determination of Nutrients in Noug Seeds

The reagents used for oil extraction, fatty acid derivatization, and gas chromatography (GC) analyses were supplied by Merck KGaA (Darmstadt, Germany) and were deemed suitable for these uses: acetic acid (glacial, 100%), methanol (anhydrous), chloroform (min 99.8%), sulphuric acid (95.0–98.0%), and hexane (≥95%). The methyl-heptadecanoate (≥99%) used as an internal standard for GC analysis of the oil was purchased from Sigma–Aldrich (Stockholm, Sweden). The nitric acid (BAKER INSTRA-ANALYZED® reagent grade) used for digesting the fine powder of noug samples for mineral content analysis was supplied by Avantor™ (Radnor, PA, USA).

2.3. Lipid Extraction, Methylation of Fatty Acids, and Gas Chromatography (GC) Analysis

The content and composition of total seed oil and fatty acids were evaluated according to Bligh and Dyer [32], with the modifications described in Geleta et al. [27]. Thus, three replications, each having 10 seeds, from each genotype and site were homogenized using 1 mL 0.15 M acetic acid and 3.75 mL MeOH:$CHCl_3$ (2:1, v/v), before adding 1.25 mL chloroform and 0.9 mL Millipore water, followed by vortexing. After centrifugation for 2 min at 3000 rpm, 50 mL of the chloroform phase of the extracted solution was transferred to a new glass tube, placed on a sand adjusted to 70 °C, and evaporated under a weak beam of liquid nitrogen. Following the complete evaporation of the chloroform solution, the fatty acids in the glass tube were methylated by adding 2 mL of 2% H_2SO_4 in methanol and incubating for 45 min at 90 °C. After the methylated solution was cooled down to room temperature, 200 nmol methyl-heptadecanoate (17:0-ME) was added to the solution as an internal standard, followed by the addition of 0.9 mL Millipore water and 2 mL hexane. The methylated solution was centrifuged for 2 min at 2000 rpm, and then 200 µL of the hexane phase containing the fatty acid methyl esters (FAMEs) was transferred to GC vials for analysis on Shimadzu GC model 17A (Kyoto, Japan), connected to a flame ionization detector [27]. The proportion of the fatty acids in the oil was calculated based on the relative percentage of the total peak area to that of the internal standard. The total lipid content of each analyzed seed sample was determined based on the total amount of the fatty acids (including the four major fatty acids; linoleic, oleic, palmitic, and stearic acids), as described in Geleta et al. [27]. This includes a calculation of the weight of each fatty acid and glycerol in the GC sample. Thereafter, the weight of triacylglycerol in the GC sample was calculated based on the weight of the fatty acids and the weight of glycerol in the sample. Then, the total lipid content was determined based on the weight of triacylglycerol and the dry weight of the seed sample [27].

2.4. Determination of the Protein Content

The total nitrogen content of each seed sample was determined on two sample replicates by dry combustion using an automated nitrogen analyzer (Model: Vario MAX CN, Elementar Americas) [33]. The protein content of each sample was then calculated from the estimated total nitrogen content (N) of the samples, using 5.6 [34] as the conversion factor.

2.5. Analysis of the Mineral Contents

A mineral content analysis of noug seeds was conducted on two sample replicates for each sample using a Perkin-Elmer Optima® 8300 inductively coupled plasma optical emission spectrometer (ICP-OES, Perkin-Elmer, Waltham, MA, USA) [35,36]. Each sample was ground to a fine powder, and 0.5 g of each ground sample was diluted with water to a volume of 50 mL prior to analysis. The diluted samples were then placed in a microwave oven for digestion in a closed vessel with 7 mL HNO_3 and 3 mL distilled water at a pressure and temperature of 375 psi and 185 °C, respectively. Blank solutions (without sample) were used as the control, and cross contaminations were checked. The mineral contents were recorded in micrograms of the mineral per gram of flour added to the solution (µg g^{-1}) for each analyzed sample.

2.6. Statistical Analysis

An evaluation of the data was carried out using the statistical package, SAS version 9.4 (SAS Institute Inc., Cary, NC, USA). Due to the unbalanced dataset, the mixed model (PROC MIXED), followed by the Tukey-Kramer test, was used to calculate the effect of the genotypes and location, as well as their interaction on the analyzed components, such as the total lipid content, content of various fatty acids, and mineral and protein content. Principal component analysis was carried out to understand the genotypic relationship for the combined variation in total lipid, fatty acids, mineral, and protein contents.

3. Results and Discussion

3.1. Quantity and Quality of Lipids

The mean lipid content of the analyzed genotypes across the three environments ranged from 32.5% (*NG-85*) to 45.7% (*NG-99*) (Table 1), with about 15% of the genotypes showing a mean lipid content of 40% or above. Thus, the lipid content of the noug genotypes in the present study was in parity with the previously reported lipid content of 27–56% in this crop [19,23–27]. Previous studies have reported a two-fold variation between genotypes [27], with higher values in cultivated noug (41–43%) than in wild *Guizotia* (21–33%) [23] and similar levels in noug produced in the USA as in noug imported from India or Ethiopia to the USA (36–38%) [19]. Mixed model results revealed that the genotype and environment, as well as their interactions (G × E), influenced the lipid content in the evaluated noug genotypes (Table 2).

Table 2. F-values for the sources of variation (i.e., environments (E), genotype (G), and G × E interactions) from the mixed model for the total oil content and fatty acids in noug.

Source of Variation	DF †	Total Lipid Content	Linoleic Acid	Oleic Acid	Palmitic Acid	Stearic Acid
Environment (E)	2	29.6 ***	32.7 ***	38.7 ***	ns	20.5 ***
Genotype (G)	44	2.7 ***	ns	ns	ns	ns
G × E	44	2.3 ***	1.4 ***	4.0 ***	1.4 *	4.1 ***

† DF = degrees of freedom, * and *** = Significant at $p < 0.05$ and 0.005, respectively, ns = not significant at $p < 0.05$.

The fatty acid composition of the noug genotypes evaluated in the present study corresponded well with previous investigations, which found linoleic, oleic, palmitic, and stearic acids to be the major fatty acids in noug seeds [23,24,26,27]. Thus, linoleic acid was the predominant fatty acid in all analyzed samples, with values ranging from 72.3% (*NG-113*) to 77.8% (*NG-93*); the oleic acid concentration ranged from 5.2% (*NG-87*) to 9.2% (*NG-113*); and palmitic and stearic acids ranged from 7.8% (*NG-113*) to 9.6% (*NG-87*) and 7.2% (*NG-111*) to 10.0% (*NG-113*), respectively (Table 1). In previous studies, linoleic acid was found to range from 32–58% [26], 54–73% [23], and above 70% [24], while oleic acid was in the range of 6–11% [24], 5.4–27% [23], 3.3–31% [27], and 23–53% [26]. The present study showed clear effects of the environment and of the interactions between the genotype and environment on the content of fatty acids (Table 2). Such relationships between the cultivation environment and fatty acids composition have not been reported previously for noug. However, the genotype and environment, as well as their interactions, are well known to influence most components in plants [37,38], and the effects on the content of total lipids and fatty acids have been previously reported, e.g., in sunflower (*Helianthus annuus* L.), which is the closest crop to noug [39–41].

Among the fatty acids, the content of oleic acid was impacted by the environment and genotype x environmental interactions to a higher degree than the other fatty acids (Table 2). Additionally, previous studies have revealed a strong environmental impact on the oleic acid content in noug, where the altitudes of production have a significant effect on certain genotypes [27,42]. Specifically, noug genotypes with a high oleic acid content, which originate in low-altitude areas of Ethiopia, have been found to continuously produce high contents of oleic acid when grown at low altitudes, although the content of oleic acid was found to decrease substantially when grown at high altitudes [27]. The genotype, *NG-84*, has been reported in previous studies as such a high oleic acid genotype when grown at low altitudes, with a decreased oleic acid content at high altitudes and high oleic content (52%) when grown at 21 °C–25 °C under green-house conditions [27]. In the present study, this genotype did not produce such a high oleic acid content, although the levels were still higher at lower altitudes: a 9.8% oleic acid content at Debre Zeit (1900 masl), 7.1% at Ginchi (2300 masl), and 4.8% at Holeta (2400 masl). However, the present study also revealed a significantly positive ($p < 0.01$) Spearman rank correlation between high oleic acid values at low altitudes and the standard deviation of oleic acid content at various

altitudes, independent of the altitude of origin of the samples (Figure 2). Thus, in this study, we were able to confirm a previous finding [27] that shows low-altitude areas of Ethiopia to be the origin for high-oleic-acid noug e.g., *NG-88* (referred to as *Tg-2* in Geleta et al. [27]), which produced 31.1% oleic acid under green-house conditions, collected at 1680 masl. This study also revealed that low-oleic-acid-content genotypes originated from low-altitude areas, showing less variation in the oleic acid content when grown at various altitudes (Figure 2).

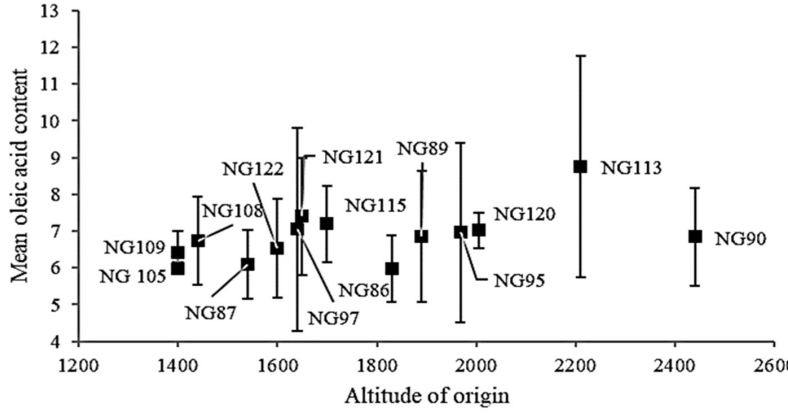


Figure 2. Variation of oleic acid content against altitude of genotype origin.

Furthermore, a high content of oleic acid, when grown at high altitudes, was rarely found among the evaluated material. Therefore, our findings indicate limited opportunities to breed for and produce high levels of oleic acid in noug grown at high altitudes, while it would be worthwhile to focus on developing high-oleic-acid noug cultivars, specifically for production in lowland areas of Ethiopia and elsewhere. Opportunities to produce high-linoleic and oleic acid noug varieties would contribute an additional asset to the highly nutritious noug oil.

3.2. Protein Concentration in Noug Seeds

The protein concentration in the noug seeds used in this study varied between 17.8% and 30.0%, depending on the genotype and location of cultivation, and therefore corresponded with previous reports of a 25–28% protein concentration in noug seeds [43], although with a wider range of content. ANOVA showed no significant (at $p < 0.05$) variation in the seed protein concentration among genotypes (Table 3) or in the interactions between genotypes and locations, although a significantly ($p < 0.05$) higher seed protein concentration was noted when noug was grown in Ginchi (27.5%), compared with Holeta (25.4%; Table 4).

Table 3. Mean squares from the analyses of variance (ANOVA) of the seed protein concentration and content of minerals in noug.

Source of Variation	DF †	Protein (10^1)	Minerals											
			B (10^1)	Ca (10^5)	Cu (10^1)	Fe (10^5)	K (10^7)	Mg (10^6)	Mn (10^4)	Na (10^2)	P (10^6)	S (10^6)	Se (10^{-3})	Zn (10^2)
Location	1	8.2 *	0.6	0.7	0.1	4.9 *	2.1 ***	5.1 **	1.5 ***	0.4	2.3	1.2 ***	4.9 ***	2.1 *
Genotype	35	0.6	1.3	1.9	4.2	0.9	0.1 **	0.6	0.00	0.7	0.8	0.0	0.2	0.2
Error	35	0.4	1.6	1.5	3.6	0.9	0.0	0.6	0.00	1.4	1.3	0.0	0.2	0.4

† DF = degrees of freedom, *, ** and *** = Significant at $p < 0.05$, 0.01 and 0.005, respectively.

Table 4. Mean values of the seed protein concentration (%) and of mineral content (μg g^{-1}) for those minerals showing a significant difference between the two experimental locations.

Location	N	Protein	Minerals						
			Fe	K	Mg	Mn	S	Se	Zn
Holeta	36	25.4 b	347 a	9254 a	4080 a	46.8 a	3253 a	0.043 a	45.0 b
Ginchi	36	27.5 a	181 b	8170 b	3547 b	18.0 b	3000 b	0.027 b	48.4 a

Numbers within a column followed by the same letter do not differ significantly ($p < 0.05$).

Most compounds, such as fatty acids, proteins, minerals, and phytochemicals in plant seeds are known to vary and be affected both by the genotype and the environment, and their interactions, and the magnitude of these effects on the content of the compounds depends on the genotypes and environments used [38,44]. In this study, the noug material was selected rather broadly to obtain genetic variation, and despite this, a limited genetic variation was found in the seed protein concentration, indicating little genetic variation in this trait for noug. Despite this, the cultivation environment had a significant effect on the trait. The effect of the environment on the grain/seed protein concentration is the result of complex processes, where both the protein accumulation and accumulation of starch and oil in the grain/seed influences the total protein concentration [45,46]. Thus, an increased protein concentration is often the result of a decrease in starch or oil in the seed/grain, which is often the result of stress conditions (heat, drought, insects, or pests), which negatively affect plant growth [38,47]. The fact that the plants were grown under rainfed conditions at both locations, and that Ginchi normally is a drier cultivation area than Holeta (the estimated average annual rainfall was 663 mm and 914 mm, respectively, as per the metrological records of the year), might be the explanation for the differences seen in the seed protein concentration. This implies that drought conditions for noug cultivation results in less dry matter, a lower yield, and a higher grain protein concentration.

3.3. Content of Mineral Elements

The mean seed mineral content of all the analyzed noug genotypes are presented in supplementary Table S1. Basically, the ANOVA analyses revealed a limited genotypic variation for all the minerals analyzed, with the exception of K (Table 3), while again the cultivation locations resulted in a significant variation for the majority of the minerals (Fe, K, Mg, Mn, S, Se, and Zn; Table 4). A higher mineral content was generally observed in the genotypes grown in Holeta, compared with Ginchi. However, the difference in mineral content between the two locations cannot be explained by a dilution effect of biomass variation, as the mineral content was not higher at Ginchi than at Holeta. Studies have shown that mineral content in cereal grains/seeds are known to be genetically determined but also highly dependent on the variation in the soil mineral content [38,48]. Previous studies on the mineral content of a range of grains/seeds have shown a wide variation in mineral content between types of grains/seeds and also between genotypes [30,36,48]. The fact that a limited variation was found between noug genotypes in mineral composition in the present study needs to be further evaluated. Here, we assume that the variation in soil mineral content of the two locations is the major determinant of the seed mineral content of noug. Thus, the differences in seed mineral content between the noug genotypes grown at the two locations was most likely caused by differences in the mineral content between the two localities. Therefore, environments with suitable soil conditions should be considered for noug production, if an increased seed mineral content is desired.

3.4. Variation in Oil Content, Fatty Acids, Seed Protein Concentration, and Minerals across Genotypes and Locations

Principal component analysis, used to depict the variation of all evaluated variables (total lipid content, fatty acid composition, seed protein concentration, and minerals) (Figure 3a), clearly revealed the effect of the growing locations on the levels of these variables (Figure 3b).

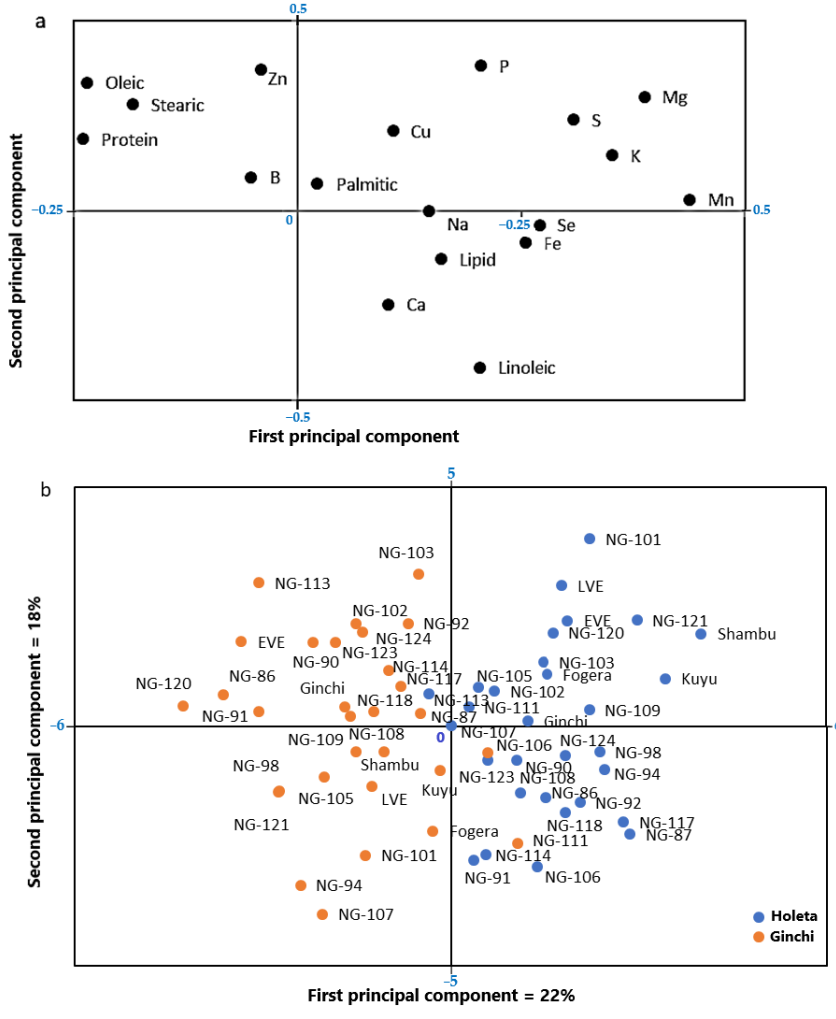


Figure 3. Loading (**a**) and score (**b**) plots from the principal component analysis of the total lipid, fatty acid, seed protein, and mineral contents of noug genotypes grown at two locations (Holeta and Ginchi).

Basically, all genotypes grown at Ginchi showed negative values for the first principal component, while the same genotypes grown at Holeta showed positive values (Figure 3b), corresponding with the abovementioned findings of the importance of the growing locations for the content and composition of the evaluated variables. The first principal component explained 22% of the variation, while 18% of the variation was explained by the second principal component. The seed protein concentration showed a negative value for the first principal component, while most of the minerals showed positive values (Figure 3a), thus verifying the higher seed protein concentration in the samples from Ginchi and higher values for the mineral content in the samples from Holeta, as also described above through the analyses of variance and mean value comparisons. No clear genotypic differences in the protein and mineral contents were shown in the principal component analysis (compare Figure 3a,b). As for the lipid and fatty acid contents, the principal compo-

nent analysis revealed the interacting effects of the genotypes and locations of cultivation, which might warrant further analyses of the stability of the contents.

3.5. Nutritional Value of Noug in Comparison with Other Oilseed Crops

The nutritional quality of noug can be divided into two parts: (i) the quality of noug oil production, which is principally determined by the total lipid content and composition of the fatty acids; and (ii) the quality of the whole seed/cake for human nutrition, which is basically determined by the fatty acid, protein, and mineral composition. The total lipid content (of importance for vegetable oil production) in noug had a close parity with that found in a range of other representative oilseed crops, all of which have recently been described in a monograph of important oil crops [49–53]. Higher values than in noug have been reported for sesame, groundnut, and castor bean, while lower values have been reported for soybean and cottonseed (Table 5).

Table 5. Content of total lipids, fatty acids, protein, Fe, and Zn in major oilseed crops seeds.

Oilseeds		Total Lipid Content (%)	Fatty Acids (%)				Protein mg/g	Fe mg/100 g	Zn mg/100 g
Common Name	Scientific Name		Linoleic Acid	Oleic Acid	Palmitic Acid	Stearic Acid		(10^1)	
Noug	*Guizotia abyssinica*	32.5–45.7 [a]	72.2–77.8 [a]	5.2–8.3 [a]	7.8–9.6 [a]	7.2–10.0 [a]	24.4–27.5 [a]	1.2–12.5 [a]	2.0–5.3 [a]
Sunflower	*Helianthus annuus*	20.5–23.9 [54]	32.2–54.3 [54]	31.9–56.9 [54]	6.6–6.8 [54]	4.0–4.1 [54]	10.0–27.1 [55]	0.5–0.7 [49]	5.0–7.6 [49]
Sesame	*Sesamum indicum*	49.5–51.3 [25]	41.0–45.0 [25]	39.5–43.0 [25]	8.4–10.3 [25]	4.5–5.8 [25]	23.1–25.2 [56]	9.3 [50]	3.8 [50]
Safflower	*Carthamus tinctorium*	36.0–41.0 [57]	74.6–78.2 [57]	11.2–14.2 [57]	6.0–6.7 [57]	2.0–2.6 [57]	17.6–18.1 [58]	3.5–4.0 [59]	1.5–2.1 [59]
Groundnut	*Arachis hypogae*	44.4–47.6 [60]	28.3–37.8 [60]	42.7–53.1 [60]	8.4–12.5 [60]	1.9–3.9 [60]	25.8 [61]	2.3 [51]	3.3 [51]
Mustard	*Brassica carinata*	39.8–46.4 [28]	17.3–19.9 [28]	10.6–12.1 [28]	3.0–3.7 [28]	1.4–2.3 [28]	32.4–36.4 [62]	1.3 [63]	0.07 [63]
Flaxseed	*Linum usitatissimum*	30.0–45.8 [25]	10.0–17.4 [25]	11.3–29.4 [25]	4.3–12.3 [25]	1.9–6.3 [25]	20.0–30.0 [64]	2.7 [52]	4.0 [52]
Soybean	*Glycine max*	20.0–22.0 [65]	50.0–60.0 [65]	22.0–25.0 [65]	7.0–10.0 [65]	2.0–5.0 [65]	37.3–40.6 [66]	7.1–8.2 [67]	4.2–11.7 [67]
Cottonseed	*Gossypium* spp.	17.5–27.0 [25]	50.5–55.0 [25]	20.0–25.0 [25]	2.0–2.5 [25]	2.5–3.3 [25]	34.2–46.3 [68]	12.0 [53]	6.1 [53]
Castor bean	*Ricinus communsis*	61.6–72.3 [69]	3.5–4.5 [69]	2.9–3.6 [69]	1.1–1.3 [69]	0.9–1.2 [69]	21–48 [70]	17.0 [71]	13.0 [71]

[a] = current study.

Considering the total lipid content, noug has the potential to develop into an oil crop of great importance. However, the nutritive value and quality of oilseeds is not only dependent on the total lipid content, but even more so on the fatty acid composition of the seed oil [72]. The fatty acid composition is also of nutritional relevance if the oil crop is used as whole seed for human consumption. Here, unsaturated fatty acids are of importance, as these are hypocholesterolemic, contributing positive health effects to the blood serum cholesterol. In this respect, linoleic and oleic acids are among unsaturated fatty acids with hypocholesterolemic properties in noug and many other oil crops, contributing positively to health when consumed [72–74]. Of specific interest here is the fact that linoleic acid is polyunsaturated and an essential fatty acid, as it cannot be produced by the human body but needs to be included in the diet [74]. The comparison of the fatty acid composition in noug with that of other oilseed crops (Table 5) revealed that noug has high levels of linoleic acid (72–78%), with only safflower showing a comparable content (74–78%). However, the content of oleic acid was generally low in noug (5–8%), and only castor bean has similar low values (3–5%), although the major fatty acid in castor bean is ricinoleic acid [75]. Despite this, the high linoleic acid content in noug may place it among the high-quality oil crops. Of relevance in this context is the storage conditions of the noug oil, which should be in airtight containers to prolong the shelf-life and to minimize the oxidation of the polyunsaturated fatty acids. Palmitic and stearic acids, also found in noug and many other oil crops, are saturated fatty acids with a hypercholesterolemic or neutral effect on the blood serum cholesterol [73]. The present comparison showed that noug has a similar level of palmitic

acid as the majority of the other oilseeds (7–10%; Table 5), but with a relatively high content of stearic acid (7–10% in relation to 1–6% in most of the other oil crops). However, the high total lipid content, combined with the high content of linolenic acid, makes noug nutritionally unique for vegetable oil production among the oilseed crops.

The protein concentration and composition are of importance if the oil crops should be used for human consumption as whole seed or seed cake. Additionally, when it comes to the seed protein concentration, noug is similar to the other oil crops. As can be seen in Table 5, only soybean and cottonseed have a higher seed protein concentration than noug, while sesame and linseed are in parity, and sunflower, safflower, and Brassica all showed lower values. The present study did not evaluate the amino acid composition of noug seed protein. Previous studies [76] have indicated that the amino acid composition of noug seeds is fairly balanced for human consumption. Protein isolates prepared from noug have shown a high in vitro protein digestibility, as well as a high iron and zinc dialysability [43]

Similarly, as for the seed protein concentration, the mineral composition is of great importance when the whole seed or cake of the oilseed crops are used for human nutrition. For mineral composition, the content of Fe in noug seeds showed a fairly high variation in the present study (Table 5). The high Fe-containing noug seeds were in parity with the high Fe content previously reported for cotton seed and castor bean [53,71]. However, the content of Zn in noug was on the lower end (2.1 to 5.1 mg 100 g^{-1}), as compared to many of the other seeds of the oilseed crops (Table 5). However, lower levels of Zn have been reported for Safflower and Mustard (Table 5) [59,63]. A sufficient intake of Fe and Zn is of relevance in most parts of the world [37], including Ethiopia. However, 10.5% to 29.6% of breastfeeding mothers in different rural areas in Ethiopia were found to have an Fe deficiency [77,78]. Furthermore, Zn deficiency, which results in stunting among children aged between 6 and 35 months was reported in both rural and urban areas in Ethiopia, as a result of a low dietary intake of Zn in the country's population in general, also including breastfeeding mothers [79]. If noug seed cake, whole-seeds, or flour should be increasingly used as a food source in Ethiopia and elsewhere, the increasing mineral content should be considered as a major breeding goal for the crop to contribute to increasing human wellness.

4. Conclusions

Noug is unique among oil crops in having a high total seed lipid content, with a high proportion of linoleic acid, contributing to the health-promoting properties of its vegetable oil. For the use of its seed cake, whole-seed, or seed meal, the fatty acid composition, with a high content of linoleic acid content, is beneficial, and noug also has a high seed protein concentration and a high Fe content, contributing to its nutritional potential. Unfortunately, the Zn content and oleic acid content are on the lower end of that found in other oilseed crops, partly hampering the quality of both the vegetable oil and whole seed/cake of noug when used for human consumption. Noug genotypes with a high oleic acid content are available when grown at low altitudes in Ethiopia, making them suitable to breed for high-oleic-acid-content noug varieties for these growing environments. However, when these genotypes are grown at high altitudes, the oleic acid content is reduced. Thus, a lack of opportunities to grow high oleic acid genotypes at high altitudes seems to prevail. The environmental conditions are also of higher importance than the genotypes in determining the grain protein concentration and mineral content in the seeds, and both factors are of importance if the oil crops are to be used as whole seed/cake for human consumption. Thus, for a high seed protein concentration, dry conditions, reducing the yield, are needed, which is not a suitable solution, while the soil mineral content is a major factor determining the mineral content of noug seeds. Despite the minor influence of the genotypes on the quality attributes, novel genomic tools, such as the recently described transcriptome-based SNP markers and the developed KASP markers, e.g., in [80], might be of importance for the future development of noug with improved nutritional properties.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/foods10081778/s1. Table S1: Mean values with standard deviations of the mineral content for the evaluated genotypes.

Author Contributions: Conceptualization, M.G., E.B., R.O., K.T.; methodology, E.J., M.G.; software, S.T., E.J.; validation, S.T., E.J., M.G.; formal analysis, S.T., E.J.; investigation, S.T., M.G.; resources, M.G.; data curation, M.G.; writing—original draft preparation, S.T.; writing—review and editing, R.O., E.J., E.B., K.T., M.G.; visualization, S.T., E.J.; supervision, R.O., E.J., M.G., E.B., K.T.; project administration, R.O., M.G.; funding acquisition, M.G., E.B., R.O., K.T., E.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Swedish Research Council (Vetenskapsrådet, VR) as part of the development research project, 2014-03517.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data presented in this study are available within the manuscript and in the supplementary material, Table S1. The raw data can be obtained by contacting the corresponding author.

Acknowledgments: The authors are grateful to the Swedish Research Council for financing this research. We would like to thank the Ethiopian Institute of Agricultural Research (EIAR) for the fruitful collaboration in providing all the necessary support and infrastructure to conduct the field trial and for providing the germplasm of the noug cultivars. We would like to thank Tomas Bryngelsson for his contribution to the acquisition of the project fund.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McKevith, B. Nutritional aspects of oilseeds. *Nutr. Bull.* **2005**, *30*, 13–26. [CrossRef]
2. Orsavova, J.; Misurcova, L.; Ambrozova, J.V.; Vicha, R.; Mlcek, J. Fatty acids composition of vegetable oils and its contribution to dietary energy intake and dependence of cardiovascular mortality on dietary intake of fatty acids. *Int. J. Mol. Sci.* **2015**, *16*, 12871–12890. [CrossRef]
3. Ayerza, R.; Coates, W. Protein content, oil content and fatty acid profiles as potential criteria to determine the origin of commercially grown chia (*Salvia hispanica* L.). *Ind. Crop Prod.* **2011**, *34*, 1366–1371. [CrossRef]
4. Amri, Z.; Lazreg-Aref, H.; Mekni, M.; El-Gharbi, S.; Dabbaghi, O.; Mechri, B.; Hammami, M. Oil characterization and lipids class composition of pomegranate seeds. *Biomed. Res. Int.* **2017**, *2017*, 2037341. [CrossRef] [PubMed]
5. El-Hamidi, M.; Zaher, F.A. Production of vegetable oils in the world and in Egypt: An overview. *Bull. Nat. Res. Cent.* **2018**, *42*, 19. [CrossRef]
6. El-Kheir, M.K.; Alamin, A.S.; Sulafa, H.N.; Sabir Ali, A.K. Composition and quality of six refined edible oils in Khartoum State, Sudan. *ARPN J. Sci. Technol.* **2012**, *2*, 177–181.
7. Asif, M. Health effects of omega-3,6,9 fatty acids: Perilla frutescens is a good example of plant oils. *Orient. Pharm. Exp. Med.* **2011**, *11*, 51–59. [CrossRef] [PubMed]
8. Teh, S.-S.; Bekhit, A. Utilization of oilseed cakes for human nutrition and health benefits. *Agric. Biomass Based Potential Mater.* **2015**, 191–229. [CrossRef]
9. Sunil, L.; Prakruthi, A.; Prasanth Kumar, P.K.; Gopala Krishna, A.G. Development of health foods from oilseed cakes. *J. Food Process. Technol.* **2016**, *7*, 11. [CrossRef]
10. Hiruy, D.; Alemseged, B. Ethnoarchaeological study of noog (*Guizotia abyssinica* (L. f.) Cass., Compositae) in Ethiopia. *Ethnoarchaeology* **2018**, *10*, 16–33. [CrossRef]
11. Marcone, M.F. Biochemical and biophysical properties of plant storage proteins, a current understanding with emphasis on 11S seed globulins. *Food. Res. Int.* **1999**, *32*, 79–92. [CrossRef]
12. Pedroche, J. Utilization of sunflower proteins. In *Sunflower*; Martínez-Force, E., Dunford, N.T., Salas, J.J., Eds.; AOCS Press: Urbana, IL, USA, 2015; pp. 395–439. [CrossRef]
13. Abiodun, O.A. The role of oilseed crops in human diet and industrial use. In *Oilseed Crops*; John Wiley & Sons Ltd.: Chichester, UK, 2017; pp. 249–263. [CrossRef]
14. Enujiugha, V.N.; Ayodele-Oni, O. Evaluation of nutrients and some anti-nutrients in lesser-known, underutilized oilseeds. *Int. J. Food Sci. Technol.* **2003**, *38*, 525–528. [CrossRef]
15. Pang, G.; Xie, J.; Chen, Q.; Hu, Z. How functional foods play critical roles in human health. *Food. Sci. Hum. Well.* **2012**, *1*, 26–60. [CrossRef]
16. Shenkin, A. Micronutrients in health and disease. *Postgrad. Med. J.* **2006**, *82*, 559–567. [CrossRef]

17. Mironeasa, S.; Codină, G.G.; Stroe, S.G. A New simple method for the evaluation of mineral elements in different oilseeds. *Commun. Soil Sci. Plant Anal.* **2016**, *47*, 1731–1737. [CrossRef]
18. Mariod, A.A.; Saeed Mirghani, M.E.; Hussein, I. Guizotia abyssinica (L.f.) Cass. Niger. In *Unconventional Oilseeds and Oil Sources*; Elsevier: San Diego, CA, USA, 2017; pp. 131–136. [CrossRef]
19. Fatima, A.; Villani, T.S.; Komar, S.; Simon, J.E.; Juliani, H.R. Quality and chemistry of niger seeds (*Guizotia abyssinica*) grown in the United States. *Ind. Crop Prod.* **2015**, *75*, 40–42. [CrossRef]
20. Agegnehu, G. Crop management practices in Noug (*Guizotia abyssinica*) in the central highlands of Ethiopia. In Proceedings of the Second Oil Crops Research and Development Review Workshop, Addis Ababa, Ethiopia, 24–26 March 2010.
21. Allaro, H. Export performance of oilseeds and ITS determinants in Ethiopia. *JCO* **2011**, *2*, 1–15. [CrossRef]
22. Alemaw, G.; Teklewold, A. An agronomic and seed-quality evaluation of noug (Guizotia abyssinica Cass.) germplasm in Ethiopia. *Plant Breed.* **1995**, *114*, 375–376. [CrossRef]
23. Dagne, K.; Jonsson, A. Oil Content and Fatty Acid Composition of Seeds of *Guizotia* Cass (Compositae). *J. Sci. Food Agric.* **1997**, *73*, 274–278. [CrossRef]
24. Dutta, P.C.; Helmersson, S.; Kebedu, E.; Alemaw, G. Variation in lipid composition of niger seed (*Guizotia abyssinica* Cass.) samples collected from different regions in Ethiopia. *J. Am. Oil Chem. Soc.* **1994**, *71*, 839–843. [CrossRef]
25. Seegeler, C.J.P. *Oil Plants in Ethiopia Their Taxonomy and Agricultural Significance*; Centref of Agricultural Publishinga and Documentation: Wageningen, The Netherlands, 1983; p. 380.
26. Yadav, S.; Kumar, S.; Hussain, Z.; Suneja, P.; Yadav, S.K.; Nizar, M.A.; Dutta, M. *Guizotia abyssinica* (L.f.) cass.: An untapped oilseed resource for the future. *Biomass Bioenergy* **2012**, *43*, 72–78. [CrossRef]
27. Geleta, M.; Stymne, S.; Bryngelsson, T. Variation and inheritance of oil content and fatty acid composition in niger (*Guizotia abyssinica*). *J. Food Compos. Anal.* **2011**, *24*, 995–1003. [CrossRef]
28. Getinet, A.; Sharma, S.M. *Niger (Guizotia abyssinica (L.F.)) Cass. Promoting the Conservation and Use of Underutilized and Neglected Crops*; Institute of Plant Genetics and Crop Plant Research, Gatersleben: Rome, Italy, 1996; pp. 1–47.
29. Gopalan, C.; Rama sastri, B.V.; Balasubramanian, S.C. *Nutritive Value of Indian Foods*; National Institute of Nutrition: Hyderabad, India, 2000; pp. 48–61.
30. Deme, T.; Haki, G.D.; Retta, N.; Woldegiorgis, A.; Geleta, M. Mineral and anti-nutritional contents of niger seed (*Guizotia abyssinica* (L.f.) Cass., linseed (*Linum usitatissimum* L.) and sesame (*Sesamum indicum* L.) varieties grown in Ethiopia. *Foods* **2017**, *6*, 27. [CrossRef] [PubMed]
31. Syume, M.; Chandravanshi, B.S. Nutrient composition of niger seed (*Guizotia abyssinica* (L. f.) Cass.) cultivated in different parts of Ethiopia. *Bull. Chem. Soc. Ethiop.* **2015**, *29*, 341–355. [CrossRef]
32. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917. [CrossRef]
33. Berner, D.L.; Brown, J. Protein nitrogen combustion method collaborative study I. Comparison with smalley total Kjeldahl nitrogen and combustion results. *JAOCS* **1994**, *71*, 1291–1293. [CrossRef]
34. Mariotti, F.; Tomé, D.; Mirand, P.P. Converting nitrogen into protein—Beyond 6.25 and Jones' factors. *Crit. Rev. Food. Sci.* **2008**, *48*, 177–184. [CrossRef]
35. Wendt, R.H.; Fassel, V.A. Induction-coupled plasma spectrometric excitation source. *Anal. Chem.* **1965**, *37*, 920–922. [CrossRef]
36. Johansson, E.; Prieto-Linde, M.L.; Larsson, H. Locally adapted and organically grown landrace and ancient spring cereals—A unique source of minerals in the human diet. *Foods* **2021**, *10*, 393. [CrossRef] [PubMed]
37. Johansson, E.; Hussain, A.; Kuktaite, R.; Andersson, S.C.; Olsson, M.E. Contribution of organically grown crops to human health. *Int. J. Environ. Res. Public Health* **2014**, *11*, 3870–3893. [CrossRef]
38. Johansson, E.; Branlard, G.; Cuniberti, M.; Flagella, Z.; Hüsken, A.; Nurit, E.; Peña, R.J.; Sissons, M.; Vazquez, D. Genotypic and environmental effects on wheat technological and nutritional quality. In *Wheat Quality for Improving Processing and Human Health*; Igrejas, G., Ikeda, T., Guzmán, C., Eds.; Springer: Cham, Switzerland, 2020; pp. 171–204.
39. Marinkovic, R.; Jockovic, M.; Marjanovic-Jeromela, A.; Jocic, S.; Ciric, M.; Balalic, I.; Sakac, Z. Genotype by environment interactions for seed yield and oil content in sunflower (*H. annuus* L.) using AMMI model. *Helia* **2011**, *34*, 79–88. [CrossRef]
40. Tabrizi, H.Z. Genotype by environment interaction and oil yield stability analysis of six sunflower cultivars in Khoy, Iran. *Adv. Environ. Biol.* **2012**, *6*, 227–231.
41. Khan, H.; Rehman, H.U.; Bakht, J.; Khan, S.A.; Hussain, I.; Khan, A.; Ali, S. Genotype × environment interaction and heritability estimates for some agronomic characters in sunflower. *J. Anim. Plant* **2013**, *23*, 1177–1184.
42. Petros, Y.; Carlsson, A.S.; Stymne, S.; Zeleke, H.; Fält, A.S.; Merker, A. Developing high oleic acid in *Guizotia abyssinica* (L.f.) Cass. by plant breeding. *Plant Breed.* **2009**, *128*, 691–695. [CrossRef]
43. Thatte, P.; Jyothi Lakshmi, A. Nutritional potential, bioaccessibility of minerals and antioxidant properties of niger (*Guizotia abyssinica* Cass.) seed protein. *Int. J. Food Sci. Technol.* **2012**, *47*, 656–663. [CrossRef]
44. Marwede, V.; Schierholt, A.; Möllers, C.; Becker, H. Genotype × environment Interactions and heritability of tocopherol contents in canola. *Crop Sci.* **2004**, *44*, 728–731. [CrossRef]
45. Jenner, C.F.; Ugalde, T.D.; Aspinall, D. The physiology of starch and protein deposition in the endosperm of wheat. *Funct. Plant Biol.* **1991**, *18*, 211–226. [CrossRef]

46. Mitchell, M.; Pritchard, J.; Okada, S.; Larroque, O.; Yulia, D.; Pettolino, F.; Szydlowski, N.; Singh, S.; Liu, Q.; Ral, J.-P. Oil Accumulation in transgenic potato tubers alters starch quality and nutritional profile. *Front. Plant Sci.* **2017**, *8*, 1–11. [CrossRef] [PubMed]
47. Dwivedi, S.; Sahrawat, K.; Upadhyaya, H.; Ortiz, R. Food, nutrition and agrobiodiversity under global climate change. *Adv. Agron.* **2013**, *120*, 1–28. [CrossRef]
48. Hussain, A.; Larsson, H.; Kuktaite, R.; Johansson, E. Mineral composition of organically grown wheat genotypes: Contribution to daily minerals intake. *Int. J. Environ. Res. Public Health* **2010**, *7*, 3442–3456. [CrossRef]
49. Khurana, S.; Singh, R. Sunflower (*Helianthus annuus*) seed. In *Oilseeds: Health Attributes and Food Applications*; Tanwar, B., Goyal, A., Eds.; Springer Nature: Singapore, 2021; pp. 123–143.
50. Sharma, L.; Saini, C.S.; Punia, S.; Nain, V.; Sandhu, K.S. Sesame (*Sesamum indicum*) seed. In *Oilseeds: Health Attributes and Food Applications*; Tanwar, B., Goyal, A., Eds.; Springer Nature: Singapore, 2021; pp. 305–330.
51. Syed, F.; Arif, S.; Ahmed, I.; Khalid, N. Groundnut (peanut) (*Arachis hypogaea*). In *Oilseeds: Health Attributes and Food Applications*; Tanwar, B., Goyal, A., Eds.; Springer Nature: Singapore, 2021; pp. 93–122.
52. Ganguly, S.; Panjagari, N.R.; Raman, R.K. Flaxseed (*Linum usitatissimum*). In *Oilseeds: Health Attributes and Food Applications*; Tanwar, B., Goyal, A., Eds.; Springer Nature: Singapore, 2021; pp. 253–283.
53. Sihag, M.K.; Patel, A.; Kumar, V. Cottonseed (*Gossypium hirsutum*). In *Oilseeds: Health Attributes and Food Applications*; Tanwar, B., Goyal, A., Eds.; Springer Nature: Singapore, 2021; pp. 73–92.
54. Kefale, B. Evaluation of basic quality parameters with in sunflower and safflower varieties from Holeta, Ethiopia. *MOJ Biorg. Org. Chem.* **2017**, *1*, 50–52. [CrossRef]
55. González-Pérez, S. 12—Sunflower proteins. In *Sunflower*; Martínez-Forcec, E., Dunford, N.T., Salas, J.J., Eds.; AOCS Press: Urbana, IL, USA, 2015; pp. 331–393. [CrossRef]
56. Bahkali, A.H.; Hussain, M.A.; Basahy, A.Y. Protein and oil composition of sesame seeds (*Sesamum indicum*, L.) grown in the Gizan area of Saudi Arabia. *Int. J. Food Sci. Technol.* **1998**, *49*, 409–414. [CrossRef]
57. Ny, A.S.; Rah, E.D. Characteristics and oxidative stability of some safflower (*Carthamus Tinctorius* L.). *J. Nutr. Food Sci.* **2015**, *S14*. [CrossRef]
58. Kutsenkova, V.S.; Nepovinnykh, N.V.; Guo, Q. Using of safflower seeds as a protein fortifier for shortbread. *Food Hydrocoll.* **2020**, *105*, 105808. [CrossRef]
59. Al Surmi, N.Y.; El Dengawy, R.A.H.; Khalifa, A.H. Chemical and nutritional aspects of some safflower seed varieties. *J. Food Process. Technol.* **2016**, *7*, 5.
60. Yusuf, Z.; Hugo, A.; Utta, H.; Mohammed, W.; Shimelis, H. Fatty acid profile of groundnut (*Arachis hypogaea* L.) cultivars in Ethiopia fatty acid profile of groundnut (*Arachis hypogaea* L.) cultivars in Ethiopia. *IJHSOP* **2019**, *5*, 89–91.
61. Arya, S.S.; Salve, A.R.; Chauhan, S. Peanuts as functional food: A review. *J. Food Sci. Technol.* **2016**, *53*, 31–41. [CrossRef] [PubMed]
62. Abul-Fadl, M.M.; El-Badry, N.; Ammar, M.S. Nutritional and chemical evaluation for two different varieties of mustard seeds. *World Appl. Sci. J.* **2011**, *15*, 1225–1233.
63. Aletor, O.; Adegoke, A.B. Evaluation of proximate chemical components, nutritional value, antinutritional and antioxidant properties in yellow mustard seeds. *Int. J. Adv. Res. Publ.* **2018**, *2*, 26–32.
64. Martinchik, A.N.; Baturin, A.K.; Zubtsov, V.V.; Molofeev, V. Nutritional value and functional properties of flaxseed. *Vopr. Pitan.* **2012**, *81*, 4–10. [PubMed]
65. Chu, W.S.; Sheldon, V.L. Soybean oil quality as influenced by planting site and variety. *J. Am. Oil Chem. Soc.* **1979**, *56*, 71–73. [CrossRef]
66. Zhu, Z.; Chen, S.; Wu, X.; Xing, C.; Yuan, J. Determination of soybean routine quality parameters using near-infrared spectroscopy. *Food. Sci. Nutr.* **2018**, *6*, 1109–1118. [CrossRef] [PubMed]
67. Szostak, B.; Glowacka, A.; Klebaniuk, R.; Kieltyka-Dadasiewicz, A. Minerla composition of traditional non-GMO soybean cultivars in relation to nitrogen fertilization. *Sci. World J.* **2020**, *2020*, 9374564. [CrossRef]
68. Bolek, Y.; Tekerek, H.; Hayat, K.; Bardak, A. Screening of cotton genotypes for protein content, oil and fatty acid composition. *J. Agric. Sci.* **2016**, *8*, 107–121. [CrossRef]
69. Bekele, B.A.; Ourgessa, A.W.; Terefe, A.A.; Hailu, S.S. Studies on Ethiopian castor seed (*Ricinus communis* L.): Extraction and characterization of lipid. *J. Nat. Prod. Resour.* **2018**, *4*, 188–190. [CrossRef]
70. Akande, T.O.; Odunsi, A.A.; Akinfala, E.O. A review of nutritional and toxicological implications of castor bean (*Ricinus communis* L.) meal in animal feeding systems. *J. Anim. Physiol. Anim. Nutr.* **2016**, *100*, 201–210. [CrossRef]
71. Adebayo, O.C.; Ogidi, C.O.; Akinyele, B.J. Nutritional value and safety of castor bean (*Ricinus communis*) seeds detoxified in solid-state fermentation by *Pleurotos ostreatus*. *Biofarmasi. J. Nat. Prod. Biochem.* **2019**, *17*, 51–60.
72. Winkler-Moser, J.K.; Mehta, B.M. Chemical composition of fat and oil products. In *Handbook of Food Chemistry*; Cheung, P.C.K., Mehta, B.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2015; pp. 365–402. [CrossRef]
73. Velasco, L.; Fernández-martínez, J.M. Breeding oilseed crops for improved oil quality. *J. Crop Prod.* **2002**, *5*, 309–344. [CrossRef]
74. Horrobin, D.F. Nutritional and medical importance of gamma-linolenic acid. *Prog. Lipid Res.* **1992**, *31*, 163–194. [CrossRef]
75. Yeboah, A.; Ying, S.; Lu, J.; Xie, Y.; Amoanimaa-Dede, H.; Boateng, K.G.A.; Chen, M.; Yin, X. Castor oil (*Ricinus communis*): A review on the chemical composition and physicochemical properties. *Food Sci. Technol.* **2020**. [CrossRef]

76. Eklund, A.; Ågren, G. Some chemical and bilólogical properties of a proteins concentrate from nigerseed (*Guizotia abyssinica* Cass.). *Qual. Plant. Mater. Vegetab.* **1970**, *19*, 313–332. [CrossRef]
77. Roba, K.T.; O'Connor, T.P.; Belachew, T.; O'Brien, N.M. Concurrent iron and zinc deficiencies in lactating mothers and their children 6-23 months of age in two agro-ecological zones of rural Ethiopia. *Eur. J. Nutr.* **2018**, *57*, 655–667. [CrossRef]
78. Harika, R.; Faber, M.; Samuel, F.; Kimiywe, J.; Mulugeta, A.; Eilander, A. Micronutrient status and dietary intake of iron, vitamin A, iodine, folate and zinc in women of reproductive age and pregnant women in Ethiopia, Kenya, Nigeria and South Africa: A systematic review of data from 2005 to 2015. *Nutrients* **2017**, *9*, 1096. [CrossRef] [PubMed]
79. Ayana, G.; Moges, T.; Samuel, A.; Asefa, T.; Eshetu, S.; Kebede, A. Dietary zinc intake and its determinants among Ethiopian children 6–35 months of age. *BMC Nutr.* **2018**, *4*, 30. [CrossRef] [PubMed]
80. Tsehay, S.; Ortiz, R.; Johansson, E.; Bekele, E.; Tesfaye, K.; Hammenhag, C.; Geleta, M. New transcriptome-based SNP markers for noug (*Guizotia abyssinica*) and their conversion to KASP markers for population genetics analyses. *Genes* **2020**, *11*, 1373. [CrossRef] [PubMed]

II

genes

MDPI

Article

New Transcriptome-Based SNP Markers for Noug (*Guizotia abyssinica*) and Their Conversion to KASP Markers for Population Genetics Analyses

Sewalem Tsehay [1,*], Rodomiro Ortiz [1], Eva Johansson [1], Endashaw Bekele [2], Kassahun Tesfaye [2,3], Cecilia Hammenhag [1] and Mulatu Geleta [1]

1 Department of Plant Breeding, Swedish University of Agricultural Sciences, P.O. Box 101, 23053 Alnarp, Sweden; rodomiro.ortiz@slu.se (R.O.); eva.johansson@slu.se (E.J.); cecilia.gustafsson@slu.se (C.H.); mulatu.geleta.dida@slu.se (M.G.)

2 Department of Microbial, Cellular and Molecular Biology, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia; endashawbw@gmail.com (E.B.); kassahuntesfaye@yahoo.com (K.T.)

3 Ethiopian Biotechnology Institute, P.O. Box 5954, Addis Ababa, Ethiopia

* Correspondence: sewalem.tsehay@slu.se; Tel.: +46-4140-5593

Received: 18 September 2020; Accepted: 18 November 2020; Published: 20 November 2020

Abstract: The development and use of genomic resources are essential for understanding the population genetics of crops for their efficient conservation and enhancement. Noug (*Guizotia abyssinica*) is an economically important oilseed crop in Ethiopia and India. The present study sought to develop new DNA markers for this crop. Transcriptome sequencing was conducted on two genotypes and 628 transcript sequences containing 959 single nucleotide polymorphisms (SNPs) were developed. A competitive allele-specific PCR (KASP) assay was developed for the SNPs and used for genotyping of 24 accessions. A total of 554 loci were successfully genotyped across the accessions, and 202 polymorphic loci were used for population genetics analyses. Polymorphism information content (PIC) of the loci varied from 0.01 to 0.37 with a mean of 0.24, and about 49% of the loci showed significant deviation from the Hardy-Weinberg equilibrium. The mean expected heterozygosity was 0.27 suggesting moderately high genetic variation within accessions. Low but significant differentiation existed among accessions (FST = 0.045, $p < 0.0001$). Landrace populations from isolated areas may have useful mutations and should be conserved and used in breeding this crop. The genomic resources developed in this study were shown to be useful for population genetics research and can also be used in, e.g., association genetics.

Keywords: genetic diversity; genotyping; *Guizotia*; Hardy-Weinberg equilibrium; heterozygosity; KASP markers; noug; population structure; SNPs; transcriptome

1. Introduction

Noug (*Guizotia abyssinica* (L. f.) Cass.) is among the cultivated species of the family Asteraceae. Similar to other *Guizotia* species, noug is diploid (2n = 30) with relatively small chromosomes [1–3]. It has a larger genome size (1C = 3.8 pg) than its closely related congeners despite having the same chromosome number [4]. For example, the genome size of *G. scabra* ssp. *schimperi*, a suggested progenitor of noug, is only 55.3% of that of noug [4]. The smaller chromosomes of *G. scabra* ssp. *schimperi* when compared with that of noug [1,3] explains the larger genome size in the latter. An increase in the noug genome over its evolutionary and domestication period is likely due to genetic mechanisms such as gene duplication [5,6] and retrotransposition [7].

Noug is an underutilized but economically important minor oilseed crop mainly cultivated in Ethiopia and India. It is considered as a semi-domesticated crop with its center of origin and diversity

in Ethiopia that was later introduced to India and other countries. Its cultivation outside Ethiopia and India covers countries such as Congo, Eritrea, Malawi, Sudan, Tanzania, Uganda, and Zimbabwe in Africa, and Bangladesh, Bhutan, and Nepal in Asia, as well as the Caribbean islands and the USA [8–11]. Noug has been cultivated mainly with low inputs, as it is recognized as a crop that can grow in poor and waterlogged soils vis-à-vis many other crops. Its ability to grow under poor agricultural management conditions makes it a good candidate for subsistence farming [10,11]. Within Ethiopia, the crop is cultivated at altitudes ranging from 1200 to 2700 m above sea level (masl) but the main cultivation areas are within the range of 1600 to 2200 masl [10,12]. The noug seed oil content can be as high as 50% and the oil is mainly composed of palmitic, stearic, oleic, and linoleic acids, with linoleic acid commonly accounting for more than 65% [12,13]. It has a significant dietary contribution as an important source of seed proteins, carbohydrates, minerals, vitamins, and fiber, in addition to its oil [10,14,15]. These nutrients are obtained through the consumption of whole seeds after being processed in various forms for their unique nutritional, cultural, and medicinal values [16]. These diverse local uses make it the most popular edible oilseed crop in Ethiopia.

Noug is the second-largest oil crop according to both harvest area and total production in Ethiopia, only surpassed by sesame [11]. However, its seed yield is lower than several other edible oil crops grown in the country. Traditional breeding efforts for improving seed yield and related traits have seen low successes, and only a few cultivars have been released [17]. In plant breeding programs, the selection of diverse germplasm possessing desirable characteristics and understanding differences between breeding materials are crucial steps towards cultivar development. Hence, the development and utilization of genome-wide markers are highly desirable to determine and manage genetic diversity within gene pools of crops. In line with this, expressed sequence tags (ESTs) and microsatellite (SSR) markers have been developed for noug [18], and a number of molecular marker-based studies have been conducted mainly to understand its population genetics [19–22]. However, molecular breeding has not been implemented in noug, mainly because available genomic resources and tools are highly limited. Publicly available genomic resources for noug that include expressed sequence tags (ESTs), complete chloroplast genome, and gene sequences used for phylogenetic studies can be found at the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/nuccore/?term=Guizotia%20abyssinica).

Single nucleotide polymorphism (SNP) markers are the most recent and popular DNA markers with a diverse use in the analyses of the genomes of crops. SNP markers are amenable to high throughput genotyping by sequencing technologies and have been extensively used for genotyping of various crops for different applications, including genetic diversity analyses, genome-wide association research, and genetic linkage mapping [23–32]. However, high-throughput SNP genotyping is less suitable and not cost-effective when the number of target SNPs is in the hundreds or less. In such cases, a relatively low-cost genotyping approach, such as the Kompetitive Allele-Specific PCR (KASP) assay is preferable. SNPs converted to KASP markers can be genotyped in small labs using fluorescent resonance energy transfer (FRET)-capable plate readers and qPCR machines. SNP markers have been successfully converted to KASP markers and used for various applications in different crops, including the faba bean [33], mung bean [34], pea [35], peanut [36], rice [37,38], rye [39], sorghum [40], and wheat [41–43].

The objectives of the present study were (1) developing new genomic resources for noug for various applications through transcriptome sequencing of noug genotypes, SNP discovery based on the transcript sequences and converting the SNPs to KASP markers, and (2) genotyping of noug accessions using the newly developed SNP/KASP markers for population genetics analyses.

2. Materials and Methods

2.1. Plant Material

Two self-compatible noug breeding lines developed from landrace populations C19 and K13 [44] were used for transcriptome sequencing. The two self-compatible lines are significantly different from one another in several characteristics, including earliness, seed shape and size, oil content, and fatty acid composition, and hence were considered as suitable germplasm for SNP discovery. For example, K13 is characterized by early maturity, larger and shinier seeds, lower oil content, and higher oleic acid content when compared with C19. Twenty-four noug accessions grown in Ethiopia (Table S1) comprising 21 landrace populations, two released cultivars, and one breeding population (developed through crossbreeding genotypes with a high oil content) were used for genotyping using newly developed SNP/KASP markers.

2.2. RNA Extraction, Transcriptome Sequencing and Assembly, and SNP Calling

The two self-compatible lines were planted in a greenhouse at the Swedish University of Agricultural Sciences (SLU), Alnarp, for RNA extraction. Total RNA was separately extracted from leaf tissue of two-week-old seedlings of the two self-compatible lines using a spectrum plant total RNA kit (Sigma-Aldrich, Stockholm, Sweden). The extracted RNA was treated by Qiagen RNase free DNase (Qiagen, Stockach, Germany) to get rid of any DNA. The RNA samples were then sent to the University of California (Davis) and library preparation, transcriptome sequencing, assembly, and SNP analysis were conducted at the Genome Sciences Center. A non-normalized cDNA library was prepared following the Illumina's guidelines and sequencing was done on the Illumina GAII, as described for the dahlia in Hodgins et al. [45]. Transcriptome assembly was done as described for Illumina sequences in Hodgins et al. [45]. A total of 4781 previously developed transcript sequences based on USDA-ARS accession PI 508077, which were functionally annotated using the Arabidopsis Information Resource (TAIR) [18] were used as a reference for SNP discovery.

After transcriptome assembly, SNP calling was made through aligning the transcript sequences to the reference sequences, and mapping the reads to the aligned sequences using BWA [46], SAMtools [47], and in-house Perl scripts. The reads of each genotype were mapped separately using the BWA aligner to generate two BAM (binary version of sequence alignment/map format) files. The SNPs between the two genotypes were determined using SAMtools. This was followed by generating a genotype table using custom-written Perl scripts in which the corresponding nucleotides of the reference transcripts are lined up with the genotype call for the two genotypes. High-quality SNPs were then obtained by filtering SNPs with at least 6× coverage. Among these, SNPs that are homozygous for both parents, having a mapped BLAST (Basic Local Alignment Search Tool) hit in the sunflower (*Helianthus annuus*), and only varying from the reference in one parent were further selected. SNPs that varied from the reference in only one parent were targeted in order to use them for genotyping of a mapping population developed using these genotypes as parents. These SNPs were further filtered by eliminating SNPs with varying sites within a 10-bp range on either side or SNPs with indels in either parent. All cases where more than one noug gene hits the same mapped *Helianthus* contig were also removed to avoid close paralogs. This filtering procedure resulted in 959 SNPs within 628 noug contigs. The accession numbers and full sequences of the 628 reference transcripts as well as the SNP sites, reference, and alternate alleles at each of 959 SNP loci are provided in Table S2.

2.3. Planting, Sampling, and DNA Extraction

The 24 noug accessions were planted in a greenhouse at SLU, Alnarp, for DNA extraction. A young leaf tissue was collected from two-week-old seedlings separately from individual plants for each accession using a LGC plant sample collection kit (KBS-9370-001) provided by LGC-Genomics (https://biosearch-cdn.azureedge.net/assetsv6/Plant-leaf-kit.pdf). Each accession was represented by 12 individuals except NG099 and NG108 (Table S1) that had 7 and 10 individuals, respectively. The 281

samples, representing the 24 accessions, collected into three deep-well plates (96-well), were then sent to LGC-Genomics (Hoddesdon, UK) where DNA extraction and genotyping were conducted. High-quality genomic DNA suitable for KASP genotyping was extracted using the sbeadex plant kit (https://www.biosearchtech.com/products/extraction-and-purification-reagents/dna-purification-kits/sbeadex-kits) at the LGC-Genomics facility.

2.4. Competitive Allele-Specific PCR (KASP) Assay Design and Genotyping

A file containing the 628 reference transcript sequences, the reference and alternate alleles of the 959 SNP loci, and their positions in the reference sequences were provided to LGC-Genomics. These data were then used for designing a competitive allele-specific PCR (KASP) assay for each SNP locus. Among the 959 SNP loci, a KASP assay was successfully designed for 931 loci, whereas 28 of them failed the KASP assay design. Hence, 931 SNP loci were used for the genotyping of the 281 samples. The genotypic data was received from LGC-genomics and analyzed using LGC's KlusterCaller software (https://www.biosearchtech.com/support/tools/genotyping-software/klustercaller) and the genotype clusters were viewed using LGC's SNP-viewer application (https://www.biosearchtech.com/support/tools/genotyping-software/snpviewer) that displays the two homozygotes and the heterozygote classes as separate clusters plate by plate.

2.5. Statistical Analyses

Various population genetics parameters were estimated using different statistical software. Genetic diversity indices were estimated for each accession or locus using Popgen32 [48], GenAlEx version 6.5 software [49], Arlequin ver 3.5 [50], and MEGA7 [51]. Pogen32 was used to calculate percent polymorphic loci and gene flow for each accession across all loci as well as the observed and effective number of alleles and allele frequency for each locus across all accessions. GenAlEx was used to calculate the Shannon information index, observed heterozygosity, standard and unbiased expected heterozygosity, and fixation indices for each accession. Arlequin was used to calculate Theta under the stepwise mutation model [52]. Tajima-Nei model [53] based estimates of average evolutionary divergence for each population were calculated using MEGA7. Arlequin was also used for the analysis of molecular variance (AMOVA) and for the Hardy-Weinberg equilibrium (HWE) test. MEGA7 was used for neighbor-joining-based cluster analysis using evolutionary distances computed based on the Tajima-Nei method [53], whereas GenAlEx was used for principal coordinate analysis (PCoA).

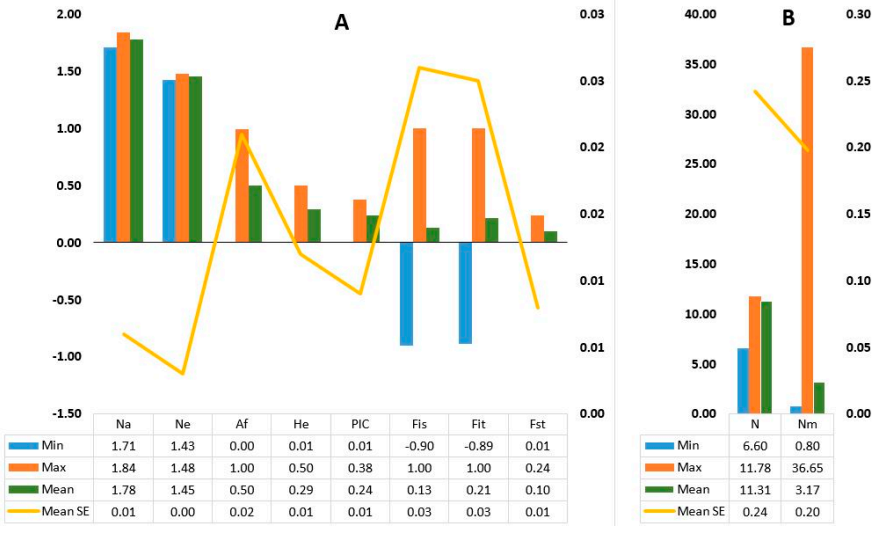
The program STRUCTURE (v. 2.3.4) [54] was used for population structure analysis using all loci and individuals. In this Bayesian approach-based analysis, an admixture model with a length of burn-in period of 100,000 and number of Markov chain Monte Carlo (MCMC) replications of 100,000 was used. The structure analysis was run for K = 1 to K = 15 with 20 runs at each K. The output of STRUCTURE was then used as input data for the STRUCTURESELECTOR program [55] for (1) determination of the optimal number of genetic clusters (K) using the ΔK method of Evanno et al. [56], and (2) graphical representation of the population genetic structure of the 24 accessions for the optimum K value through the application of the integrated CLUMPAK program of Kopelman et al. [57].

3. Results

3.1. The KASP Genotyping and the SNP Loci

Genotypic data across 202 SNP loci (Table S2) were used for population genetics analyses of the 24 accessions. Among the 931 SNPs targeted in the KASP genotyping, 554 were successfully genotyped across the 24 accessions, while 377 failed. Hence, the success rate of the KASP genotyping in this study is 59.5%. Of the 554 loci successfully genotyped, 286 (51.6%) were monomorphic across all individuals, whereas 268 (48.4%) were polymorphic. Among the 268 polymorphic loci, 66 of them (24.6%) had >10% missing data and were not used for further data analysis.

The effective number of alleles across the 202 loci varied from 1.43 to 1.48 with a mean of 1.45. Expected heterozygosity (He) varied from 0.01 to 0.50 with a mean of 0.29 whereas polymorphism information content (PIC) varied from 0.01 to 0.38 with a mean of 0.24 (Figure 1; Table S3). A large variation was observed in fixation indices among the SNP loci. The minimum, maximum, and mean values for F_{IS} were −0.90, 1.00, and 0.13, for F_{IT} were −0.89, 1.00, and 0.21, and for F_{ST} were 0.01, 0.24, and 0.10, respectively (Figure 1A). The estimate of gene flow (Nm) for each locus showed wide variation, ranging from 0.80 to 36.65, with a mean of 3.17 (Figure 1B; Table S3).



	Na	Ne	Af	He	PIC	Fis	Fit	Fst
Min	1.71	1.43	0.00	0.01	0.01	-0.90	-0.89	0.01
Max	1.84	1.48	1.00	0.50	0.38	1.00	1.00	0.24
Mean	1.78	1.45	0.50	0.29	0.24	0.13	0.21	0.10
Mean SE	0.01	0.00	0.02	0.01	0.01	0.03	0.03	0.01

	N	Nm
Min	6.60	0.80
Max	11.78	36.65
Mean	11.31	3.17
Mean SE	0.24	0.20

Figure 1. Maximum, minimum, and mean values of (**A**) observed number of alleles (Na), effective number of alleles (Ne), allele frequency (Af), expected heterozygosity (He), polymorphism information content (PIC), fixation indices (F_{IS}, F_{IT} and F_{ST}) per locus; (**B**) sample size (N) per population and estimate of gene-flow (Nm) per locus.

The Hardy-Weinberg equilibrium (HWE) test revealed that 50.7% of the loci are at HWE whereas 49.3% of loci showed significant deviation from HWE (Figure 2; Table S3). A total of 43.4% of the loci showed heterozygote deficiency with 33.5% and 9.9% showing highly significant ($p < 0.01$) and significant ($0.01 < p < 0.05$) deviation, respectively. On the other hand, 5.9% of the loci showed excess heterozygosity with 4.4% and 1.5% showing highly significant ($p < 0.01$) and significant ($0.01 < p < 0.05$) deviation, respectively. Examples of the SNP loci showing a highly significant deviation from HWE and the description of their corresponding sunflower homologs are provided in Table 1. Twelve and four of them represent loci with heterozygote excess and deficiency, respectively. Interestingly, 10 of the 12 loci that showed heterozygote excess lacked one of the three possible genotypes expected in a bi-allelic polymorphic locus under the assumption of HWE. Similarly, all four loci with heterozygote excess lacked one of the three possible genotypes. The minor allele frequency (MAF) of the four and 12 loci ranged from 0.185 to 0.470 and 0.120 to 0.470, respectively. Among these 16 SNP loci, the change in amino acid sequences of the corresponding genes was obtained in only one locus (locus 3143A). The other 15 SNPs are synonymous substitutions (Table 1). In the case of the non-synonymous substitution, the SNP resulted in a Serine/Arginine exchange (Table 1).

Table 1. Description of polymorphic single nucleotide polymorphic (SNP) loci that showed highly significant deviation from Hardy-Weinberg Equilibrium (HWE), and corresponding sunflower (*Helianthus annuus*) homologs of noug genes harboring the SNPs.

Noug Contig	SNP Locus	SNP PINC [c]	Ref_Alt [d,e]	Missing Genotype	MAF [f]	*Helianthus annuus* Homologue: Accession Number and Aligned Region
CL3143Contig1	3143A [a]	376	G_C	GG	0.131	17.8 kDa class I heat shock protein-like_LOC110904834: XM_022150704; 212..484
	3143B [a]	388	C_A	AA	0.120	
CCHT13019.b1F16.ab1	13,019A [a]	30	C_A	AA	0.389	Two-component response regulator-like PRR73_LOC110868813: XM_022118076; 1133..1903
	13,019C [b]	372	T_A	AT	0.448	
CCHT3719.b1M18.ab1	3719A [a]	502	A_G	AA	0.239	TPR2-like protein_LOC110930065: XM_022173292; 2716..3466
CCHT4593.b1B22.ab1	4593B [a]	274	A_C	None	0.389	Calnexin homolog_LOC110865890: XM_022115225; 1115..1583
CCHT4736.b1P07.ab1	4736 [a]	498	G_C	GG	0.293	Cytochrome P450 CYP82D47-like_LOC110879526: XM_022127994; 1210..1778
CCHT7954.b1C22.ab1	7954A [a]	387	T_C	CC	0.371	Uncharacterized protein_LOC110878690: XM_022127042; 145..855
CCHT8585.b1B11.ab1	8585 [a]	280	C_T	CC	0.149	Uncharacterized protein_ LOC110927216: XM_022170855; 1193..1751
CCHT10160.b1P19.ab1	10,160 [a]	328	G_C	GG	0.441	UDP-arabinopyranose mutase 1-like_ LOC110915474: XM_022160180; 62..679
CCHT13180.b1H07.ab1	13,180 [a]	460	G_A	None	0.432	40S ribosomal protein S13_LOC110937753: XM_022180210; 4..544
CCHT17789.b1J07.ab1	17,789 [a]	698	C_A	CC	0.245	Uncharacterized 38.1 kDa protein-like_LOC110867109: XM_022116245; 208..521
CCHT20996.b1G18.ab1	20,996B [a]	636	C_T	CC	0.470	Heat shock 70 kDa protein 14 like_LOC110894223: XM_022141417, 93..831
CCHT17807.b1N11.ab1	17,807 [b]	350	T_G	GT	0.250	Nucleobase-ascorbate transporter 6-like_LOC110941756: XM_022183418; 123..833
CCHT17571.b1F02.ab1	17,571 [b]	408	T_A	AT	0.470	Probable ADP-ribosylation factor GTPase-activating protein AGD14_LOC110886873: XM_022134739; 92..748
CCHT4779.b1F19.ab1	4779 [b]	255	A_C	AC	0.185	Probable E3 ubiquitin-protein ligase ARI1_LOC110929292: XM_022172423; 1130..1867

[a] = heterozygote excess; [b] = heterozygote deficit. [c] SNP PINC = SNP position in noug contig. [d] Ref_Alt = Reference allele_Alternate allele; [e] All SNPs resulted in synonymous amino acid substitution except SNP at locus 3143A that led to the non-synonymous substitution of Serine vs. Arginine, at position 99 of the amino acid sequence of the 17.8 kDa class I heat shock protein-like protein; [f] MAF = minor allele frequency.

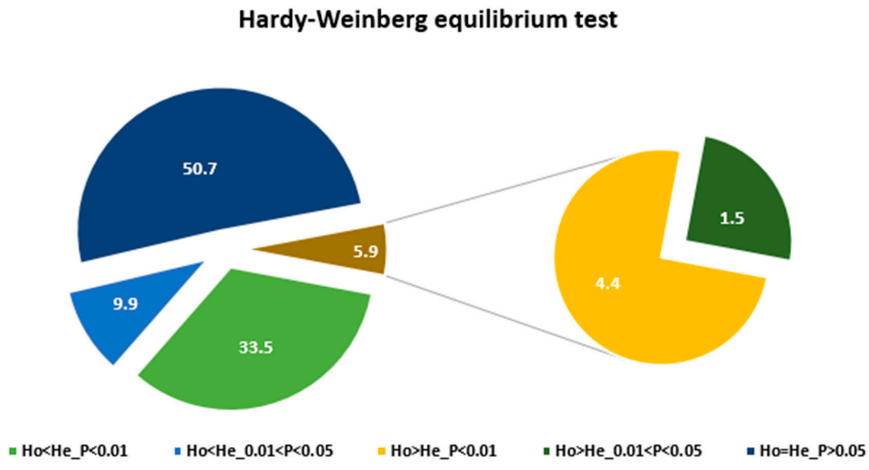


Figure 2. Pie-chart of the SNP loci significantly deviated from Hardy-Weinberg equilibrium (HWE) showing their proportions in terms of heterozygote excess and deficiency at different levels of significance.

3.2. Genetic Diversity and Population Structure

Various genetic diversity parameters were estimated for each accession based on the 202 polymorphic loci. The percent polymorphic loci (PPL) of the accessions ranged from 72% (NG108) to 85% (NG092) with a mean of 79%. The Shannon diversity index (I) ranged from 0.38 (NG124) to 0.42 (NG092) with a mean of 0.40. The lowest and highest observed (Ho) heterozygosity values were recorded for accessions NG124 (0.22) and NG099 (0.26) with a mean of 0.25. The expected heterozygosity (He) and unbiased expected heterozygosity (uHe) of the accessions ranged from 0.26 and 0.27 (NG106) to 0.28 and 0.29 (NG092) with a mean of 0.27 and 0.28, respectively. The fixation index (F) of the accessions ranged from 0.01 (NG099) to 0.16 (NG092) (Table 2). There was low variation in the Theta (H) estimated from mean heterozygosity under the stepwise mutation model [52] among the accessions with the values ranging from 1.93 (NG092) to 2.08 (NG103). The Tajima-Nei model [53] based estimates of average evolutionary divergence over sequence pairs within accessions were calculated using (1) all polymorphic loci (EAED1), (2) only polymorphic loci with an allele frequency ranging from 0.3 to 0.7 (EAED2), and (3) only polymorphic loci with minor allele frequency (MAF) of less than 0.3 (EAED3) (Table 2). The lowest values of EAED1 (0.22), EAED2 (0.38), and EAED3 (0.15) were recorded for accession NG103. The corresponding highest values were 0.31 (for accession NG086), 0.57 (for accession NG111), and 0.24 (for accession NG109), in that order. The mean values for Theta, EAED1, EAED2, and EAED3 were 2.00, 0.28, 0.48, and 0.200, respectively (Table 2).

Table 2. Summary of genetic diversity estimates for 24 noug accessions based on 202 single-nucleotide polymorphic (SNP) loci or its subsets, and mean values for different groups and all accessions.

Acc	PPL	I	Ho	He	uHe	F	Theta	$EAED_1$	$EAED_2$	$EAED_3$
NG086 [a]	75.7	0.391	0.238	0.262	0.274	0.105	2.016	0.313	0.554	0.207
NG088 [a]	77.7	0.396	0.242	0.263	0.276	0.092	2.004	0.276	0.468	0.198
NG089 [a]	75.7	0.387	0.248	0.258	0.270	0.042	2.038	0.294	0.528	0.201
NG090 [a]	80.1	0.403	0.244	0.268	0.280	0.084	1.978	0.274	0.489	0.187
NG092 [a]	84.0	0.419	0.235	0.278	0.290	0.165	1.927	0.304	0.541	0.206
NG095 [a]	80.6	0.415	0.256	0.277	0.289	0.087	1.934	0.285	0.497	0.196
NG096 [a]	73.8	0.381	0.259	0.255	0.267	0.017	2.057	0.257	0.454	0.176
NG097 [a]	76.7	0.384	0.224	0.254	0.266	0.122	2.063	0.256	0.430	0.183
NG098 [a]	78.6	0.403	0.256	0.267	0.279	0.060	1.984	0.280	0.436	0.216
NG099 [a]	76.7	0.400	0.261	0.266	0.288	0.011	1.968	n/c	n/c	n/c
NG101 [a]	79.1	0.405	0.259	0.270	0.282	0.047	2.012	0.248	0.396	0.187
NG103 [a]	80.1	0.396	0.245	0.263	0.275	0.067	2.076	0.218	0.380	0.146
NG105 [a]	78.6	0.395	0.259	0.263	0.274	0.057	1.971	0.249	0.386	0.195
NG106 [a]	77.2	0.382	0.227	0.252	0.264	0.103	2.067	0.286	0.548	0.188
NG107 [a]	75.7	0.402	0.246	0.270	0.282	0.086	1.959	0.286	0.496	0.201
NG108 [a]	70.9	0.380	0.237	0.255	0.269	0.081	1.974	0.272	0.484	0.190
NG109 [a]	79.6	0.408	0.244	0.272	0.284	0.115	2.031	0.306	0.465	0.242
NG111 [a]	80.1	0.405	0.234	0.269	0.281	0.134	2.045	0.278	0.568	0.165
NG112 [a]	79.1	0.392	0.233	0.259	0.271	0.110	1.940	0.295	0.532	0.203
NG113 [a]	77.7	0.388	0.237	0.257	0.269	0.073	2.073	0.274	0.453	0.203
NG123 [a]	82.5	0.414	0.256	0.275	0.288	0.090	2.008	0.283	0.462	0.214
NG124 [b]	72.8	0.379	0.217	0.253	0.264	0.133	1.936	0.287	0.493	0.204
Fogera [c]	80.1	0.407	0.253	0.270	0.283	0.072	1.969	0.300	0.519	0.213
Shambu [c]	81.1	0.398	0.238	0.262	0.274	0.091	2.011	0.284	0.487	0.205
Mean_Alt-1	77.8	0.397	0.245	0.264	0.277	0.088	2.005	0.275	0.447	0.205
Mean_Alt-2	77.5	0.397	0.244	0.265	0.277	0.086	2.000	0.289	0.517	0.196
Mean_Alt-3	79.0	0.398	0.245	0.265	0.278	0.075	2.013	0.265	0.470	0.182
Mean_Reg-1	78.8	0.401	0.251	0.266	0.279	0.076	2.006	0.281	0.458	0.211
Mean_Reg-2	73.8	0.382	0.240	0.255	0.267	0.073	2.031	0.262	0.456	0.183
Mean_Reg-3	79.7	0.405	0.244	0.270	0.282	0.100	1.983	0.285	0.495	0.197
Mean_Reg-4	77.4	0.391	0.236	0.260	0.272	0.093	2.010	0.285	0.507	0.199
Mean_Reg-5	80.1	0.396	0.245	0.263	0.275	0.067	2.076	0.218	0.380	0.146
Mean_Reg-6	79.1	0.407	0.250	0.271	0.286	0.077	1.982	0.282	0.533	0.181
Mean_Landrace	78.1	0.397	0.245	0.264	0.277	0.083	2.006	0.277	0.478	0.195
Mean_Cultivar	80.6	0.4025	0.245	0.266	0.278	0.081	1.99	0.292	0.503	0.209
Mean_all	78.1	0.397	0.244	0.264	0.277	0.085	2.002	0.278	0.481	0.197
SE_all	0.006	0.004	0.003	0.003	0.003	0.006	0.050	0.005	0.011	0.004

Acc = accessions ([a] landrace populations, [b] breeding population, [c] cultivars); PPL = percent polymorphic loci; I = Shannon's information index; Ho = observed heterozygosity; He = expected heterozygosity; uHe = unbiased expected heterozygosity; F = fixation index; Theta (H) = Theta from mean heterozygosity under the stepwise mutation model [52]; EAED_1, 2, and 3 = estimates of average evolutionary divergence over sequence pairs within populations for (a) all polymorphic loci, (b) loci with allele frequency of equal or above 0.3 and below or equal 0.7, and (c) loci with allele frequency of below 0.3 and above 0.7, respectively, as estimated based on the Tajima-Nei model [53]. Note: In all cases, genotypes with more than 7% missing data were excluded. The Pearson correlation coefficient of $EAED_1$ vs. $EAED_2$, $EAED_1$ vs. $EAED_3$, and $EAED_2$ vs. $EAED_3$ were 0.79 ($p < 0.001$), 0.75 ($p < 0.001$), and 0.20 ($p = 0.35$). The 21 landrace accessions were grouped into three altitudinal groups and six regional groups (see Table S1).

The analysis of molecular variance (AMOVA) was conducted without grouping the accessions as well as by grouping them according to their geographic region of origin or altitudinal range of their collection sites (Table 3). The analysis showed that 95.5% of the total variation accounted for variation within accessions whereas 4.5% accounted for the variation among them ($F_{ST} = 0.045$, $p < 0.0001$). The vast majority of the within accession variation (93.9%) was attributed to the variation within individuals (heterozygosity). Hierarchical AMOVA was conducted by grouping 21 of the 24 accessions (excluding the two cultivars and the breeding population) into (1) three altitudinal groups, (2) six geographical regions (regions-I), and (3) two geographical regions (regions-II) (Table 3). However, only 0.12% of the total variation accounted for the variation among the altitudinal groups, which is not

significant (F_{CT} = 0.001 and p = 0.19). Similarly, there was no significant differentiation between the six regions-I groups (F_{CT} = −0.001 and p = 0.67). However, there was significant differentiation between the two regions-II groups (Oromia vs. Amhara-Tigray) (F_{CT} = 0.002 and p = 0.047).

Table 3. Analysis of molecular variance (AMOVA), based on 1023 permutations, for 24 accessions without grouping, and for 21 accessions by grouping them according to altitudinal range or regions of origin.

Source of Variation	DF	Sum of Squares	Variance Components	Percentage of Variation	Fixation Indices	Probability (p) Value
Among accessions	23	1078.2	1.019 Va	4.52	F_{ST} = 0.045	Va & F_{ST} < 0.0001
Among individuals within accessions	259	5921.0	1.321 Vb	5.86	F_{IS} = 0.061	Vb & F_{IS} < 0.0001
Within individuals	283	5722.0	20.219 Vc	89.63	F_{IT} = 0.104	Vc & F_{IT} < 0.0001
Total	565	12,721.3	22.556			
[a] Among alt groups	2	97.9	0.026 Va	0.12	F_{CT} = 0.001	Va & F_{CT} = 0.1935
Among accessions within alt groups	18	800.1	1.015 Vb	4.67	F_{SC} = 0.047	Vb & F_{SC} < 0.0001
Within accessions	471	9746.7	20.694 Vc	95.21	F_{ST} = 0.048	Vc & F_{ST} < 0.0001
Total	491	10,644.8	21.735			
[b] Among regions-I	5	217.0	−0.027 Va	−0.12	F_{CT} = −0.001	Va & F_{CT} = 0.6715
Among accessions within regions-I	15	681.0	1.056 Vb	4.86	F_{SC} = 0.048	Vb & F_{SC} < 0.0001
Within accessions	471	9746.7	20.694 Vc	95.26	F_{ST} = 0.047	Vc & F_{ST} < 0.0001
Total	491	10,644.8	21.723			
[c] Among regions-II	1	55.4	0.044 Va	0.20	F_{CT} = 0.002	Va & F_{CT} = 0.047
Among accessions within regions-II	19	842.6	1.011 Vb	4.65	F_{SC} = 0.047	Vb & F_{SC} < 0.0001
Within accessions	471	9746.7	20.694 Vc	95.15	F_{ST} = 0.048	Vc & F_{ST} < 0.0001
Total	491	10,644.8	21.723			

Note: 21 of the 24 accessions were grouped according to region or altitudinal (alt) range or origin. The two cultivars and the breeding population were excluded from the grouping, as they cannot be placed in any of the groups. [a] The 21 accessions were grouped into three altitudinal groups: 1400–1680 m above sea level (masl), 1820–1968 masl, and 2045–2590 masl. [b] The 21 accessions were grouped into six regions (regions-I), and [c] the 21 accessions were grouped into two regions (regions-II) (Table S1).

AMOVA-based F_{ST} was also computed for each pair of the 24 accessions, and 268 of the 276 pairs (97.1%) showed significant differentiation between them with F_{ST} values ranging from 0.017 to 0.124 (Table 4). Only eight pairs failed to show significant differentiation (F_{ST} ranging from 0.006 to 0.012). The lowest and the highest F_{ST} values were recorded for NG111 vs. NG123 and NG096 vs. NG097, respectively (Table 4). The mean F_{ST} values that reflect the average differentiation of each accession from all other accessions ranged from 0.028 (Shambu) to 0.076 (NG097). Accessions NG096 and NG124 also showed higher differentiation from the other accessions having mean F_{ST} values of 0.074 and 0.072, respectively. On the other hand, Fogera and NG123 are among the least differentiated accessions with mean F_{ST} values of 0.031 and 0.033, respectively.

Table 4. Analysis of molecular variance (AMOVA)-based pairwise F_{ST} between the 24 populations with 1023 permutations (below the diagonal) and mean F_{ST} of each accession against all other accessions (diagonal).

Acc	086	088	089	090	092	095	096	097	098	099	101	103	105	106	107	108	109	111	112	113	123	124	Fog	Sha
086	**0.036**																							
088	0.031	**0.041**																						
089	0.026	0.041	**0.045**																					
090	0.031	0.055	0.041	**0.052**																				
092	0.023	0.026	0.046	0.041	**0.038**																			
095	0.047	0.031	0.051	0.056	0.043	**0.052**																		
096	0.054	0.075	0.076	0.081	0.061	0.095	**0.074**																	
097	0.069	0.066	0.065	0.082	0.065	0.064	0.124	**0.076**																
098	0.046	0.061	0.037	0.073	0.037	0.074	0.079	0.08	**0.055**															
099	0.030	0.053	0.042	0.053	0.040	0.048	0.078	0.082	0.066	**0.048**														
101	0.034	0.032	0.053	0.065	0.050	0.064	0.08	0.078	0.057	0.075	**0.052**													
103	0.024	0.031	0.031	0.031	0.017	0.041	0.075	0.069	0.052	0.027	0.036	**0.035**												
105	0.034	0.033	0.043	0.045	0.042	0.046	0.069	0.084	0.049	0.054	0.048	0.038	**0.047**											
106	0.031	0.036	0.039	0.044	0.036	0.049	0.072	0.069	0.043	0.054	0.042	0.024	0.035	**0.043**										
107	0.023	0.039	0.039	0.033	0.033	0.043	0.069	0.078	0.047	0.034	0.042	0.025	0.035	0.035	**0.039**									
108	0.038	0.057	0.059	0.057	0.031	0.056	0.103	0.091	0.042	0.068	0.068	0.043	0.051	0.043	0.038	**0.056**								
109	0.050	0.037	0.050	0.061	0.040	0.05	0.072	0.064	0.048	0.059	0.047	0.029	0.056	0.05	0.054	0.068	**0.052**							
111	0.025	0.029	0.028	0.041	0.027	0.042	0.05	0.086	0.056	0.025	0.045	0.024	0.037	0.045	0.034	0.055	0.042	**0.039**						
112	0.034	0.046	0.053	0.049	0.039	0.045	0.072	0.089	0.05	0.037	0.057	0.047	0.041	0.046	0.031	0.063	0.06	0.043	**0.049**					
113	0.038	0.059	0.053	0.041	0.050	0.057	0.066	0.075	0.072	0.023	0.061	0.03	0.062	0.055	0.033	0.07	0.061	0.046	0.041	**0.051**				
123	0.026	0.007 *	0.020	0.039	0.018	0.044	0.057	0.055	0.034	0.037	0.027	0.019	0.021	0.025	0.023	0.043	0.047	0.006 *	0.037	0.043	**0.031**			
124	0.065	0.056	0.077	0.091	0.068	0.069	0.09	0.083	0.086	0.062	0.078	0.072	0.084	0.074	0.065	0.093	0.076	0.06	0.083	0.065	0.058	**0.072**		
Fog	0.028	0.019	0.040	0.042	0.025	0.033	0.055	0.058	0.038	0.031	0.026	0.021	0.046	0.032	0.031	0.038	0.043	0.028	0.031	0.035	0.011 *	0.038	**0.033**	
Sha	0.012 *	0.023	0.023	0.036	0.019	0.04	0.054	0.065	0.03	0.034	0.031	0.007 *	0.028	0.020	0.020	0.023	0.021	0.011 *	0.03	0.032	0.010 *	0.064	0.010 *	**0.028**

* = No significant differentiation between the pair of accessions. The first column and row are accession names without their two initial letters (*NG*) for the first 22 accessions. Acc = Accession; Fog = Fogera and Sha = Shambu. Bold: mean F_{ST} values of each accession against all other accessions.

One hundred twenty-six individuals having genotypic data for all loci (no missing values) were selected across the 24 accessions and the evolutionary distance between each pair of individuals was computed for loci with MAF below 0.3 using Tajima-Nei method [53]. The neighbor-joining cluster analysis based on the evolutionary distances between the individuals resulted in three major clusters and several sub-clusters (Figure 3). Except in a few cases, individuals from the same accessions were placed in more than one cluster. For example, accession NG105 was represented by eight individuals, of which three, three, and two individuals were placed in cluster-I, II, and III, respectively. On the other hand, all six individuals of Fogera, a cultivar, were clustered in cluster-I and all four individuals of NG111 were placed in cluster-III. The Tajima-Nei method [53] based evolutionary distance was also used for neighbor-joining cluster analysis at the accession level using three sets of loci: all polymorphic loci, only loci with MAF of <0.3, and only loci with MAF of ≥0.3 (Figure 4). As depicted in Figure 4A–C, the clustering patterns of the accessions are clearly different among the three data sets. The analysis revealed that the clustering patterns of the accessions according to their geographic regions or altitudinal range of origin were poorly defined.

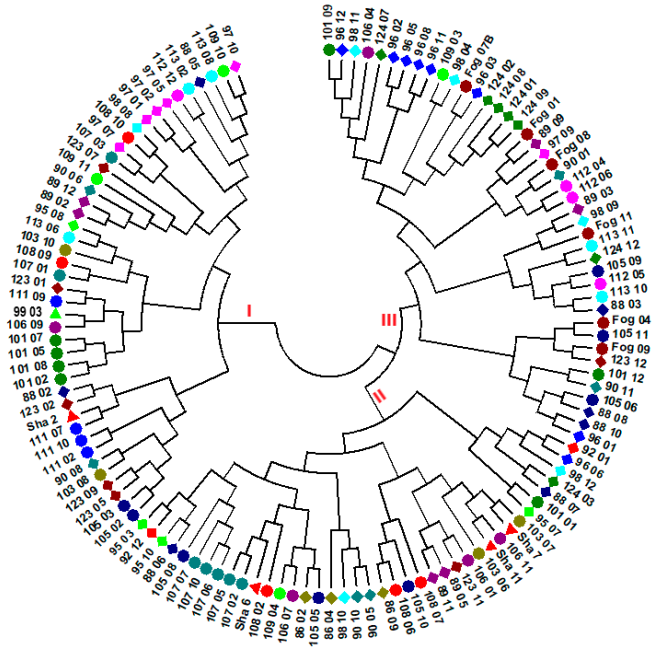


Figure 3. Neighbor-joining tree of 126 individuals representing the 24 accessions generated based on loci with a minor allele frequency of <0.3, using evolutionary distances computed by the Tajima-Nei method (Tajima and Nei 1984). The individual samples were coded in a way that the first two or three digits/letters represent their accessions and the last two-digit numbers represent the codes for the plant in that accession. The accession names are given without the two initial letters (NG), and Fog and Sha represent Fogera and Shambu, respectively. Individuals represented by the same shape and color belong to the same accession.

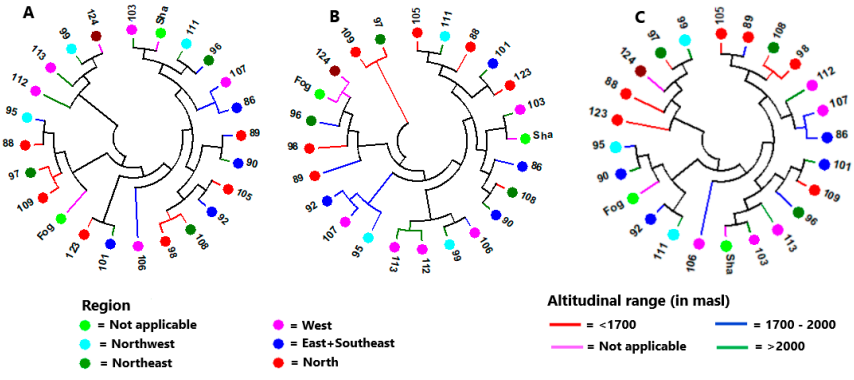


Figure 4. Neighbor-joining tree of the 24 accessions generated based on (**A**) all loci, (**B**) loci with a minor allele frequency of <0.3, and (**C**) loci with a minor allele frequency of ≥0.3 using evolutionary distances computed by the Tajima-Nei method [53]. The accession names are given without the two initial letters (NG) for 22 of the 24 accessions, and "Fog" and "Sha" represent Fogera and Shambu, respectively. Accessions represented by the same shape and color belong to the same region, and accessions represented with the same color tree-line belong to the same altitudinal range.

Principal coordinate analysis (PCoA) was also conducted to determine the relationship between the accessions. In the two-dimensional plot generated, the first and the second coordinates explained 24% and 20% of the total variation among the accessions (Figure 5A). This analysis revealed that most of the accessions were tightly clustered together suggesting low differentiation among them. However, accessions NG097, NG124, and NG096 were clearly separated from the other accessions in this two-dimensional plot.

The admixture model-based population genetic structure analysis conducted using STRUCTURE [54] and STRUCTURESELECTOR programs [55] revealed that the optimal number of genetic clusters (K) is three, as per the ΔK method of Evanno et al. [56] (Figure 5B). Hence, the 281 individuals representing the 24 accessions were reduced to three genetic populations. The graphical representation of the population genetic structure of the 24 accessions at K = 3, clearly showed that all accessions have alleles that originated from the three clusters (genetic populations), thus suggesting low differentiation between the accessions. In line with the results of the PCoA, accessions NG096, NG097, and NG124 showed higher differentiation than the other accessions in this analysis. The alleles of NG096, NG097, and NG124 were predominantly originated from cluster-2 (orange), cluster-3 (purple), and cluster-1 (blue), respectively (Figure 5C).

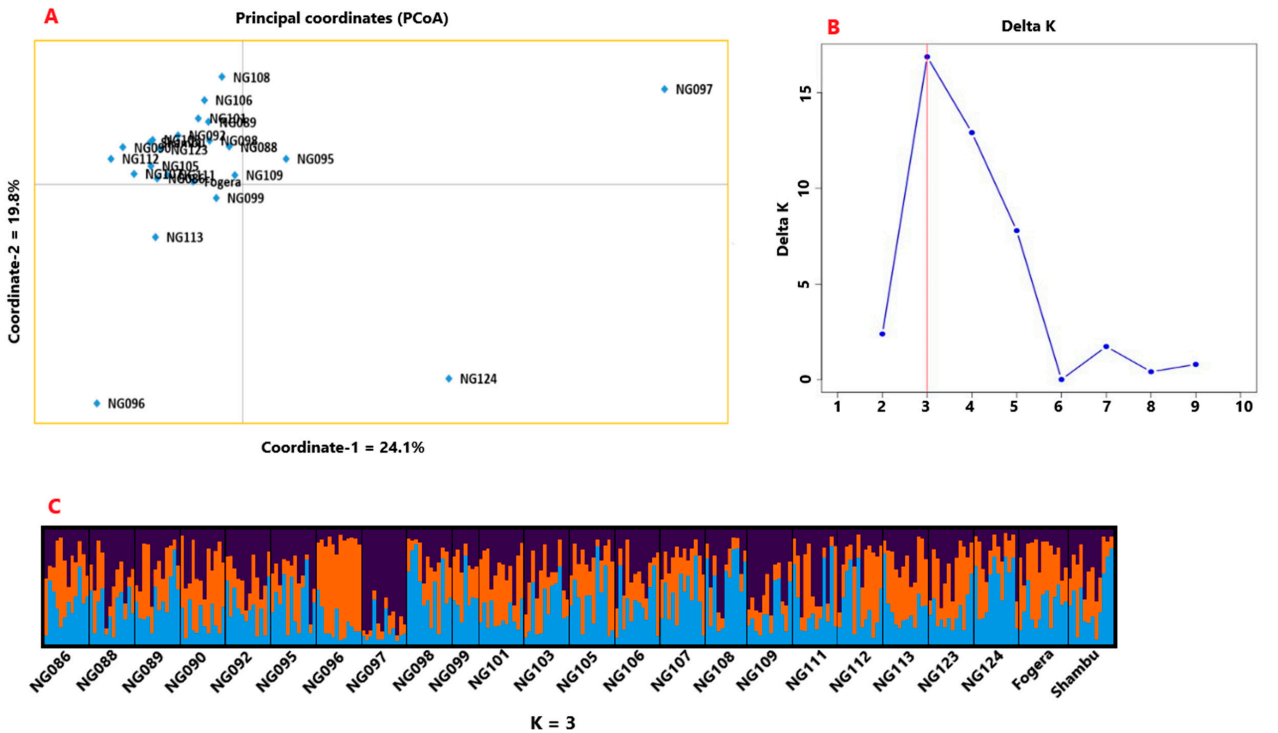


Figure 5. (**A**) Principal coordinate analysis (PCoA) based two-dimensional plot for the 24 accessions, in which the first and the second axes explained 24% and 20% of the total variation, respectively, (**B**) ΔK plot showing its maximum value at K = 3 suggesting the optimal number of genetic clusters (populations) of three, and (**C**) graphical representation of the population genetic structure of the 24 accessions for K = 3. The three colors represent the three clusters and the proportion of each color in each accession represents the average proportion of the alleles that placed each accession under the three clusters.

4. Discussion

The present study used transcriptome sequencing for the development of novel SNP markers for noug, which were used to design KASP assays and for genotyping 24 Ethiopian noug accessions. The fact that transcript sequences of only two genotypes were used for SNP discovery and the multi-step stringent procedures followed to identify and filter the SNP markers led to a relatively small number of SNPs (959). The use of a few genotypes as an SNP discovery panel can lead to an ascertainment bias [58,59]. This means that our SNP discovery approach may have left rare alleles in noug gene pool corresponding to the target transcripts undiscovered. However, the 202 polymorphic loci used in this study included loci with minor allele frequency (MAF) ranging from below 1% (rare alleles) to almost 50%, and hence the SNP discovery approach is not expected to affect the results of population genetics analyses. Although 97% of these SNPs passed the KASP assay design, only 59.5% of them were successfully used for genotyping the 24 accessions. In other words, 377 SNPs that passed the assay design (40.5%) failed at the genotyping stage. The failure could be because the SNPs do not exist in the germplasm targeted for genotyping or the primers failed to anneal to the target sequences due to sequence variation. In this study, only bi-allelic SNPs developed based on the two genotypes were used. If such SNPs are restricted to a small subset of the crop's gene pool, it is highly likely that they fail (including the case when a nucleotide at SNP locus is different from both alleles) when applied to a diverse germplasm. Further research through resequencing of the target regions using Sanger sequencing or targeted genotyping by sequencing methods, such as SeqSNP (LGC Genomics) will shed light on the factors behind the failure.

Genetic polymorphism research remains key for the understanding of variability in organisms' genetic makeup, and it is usually a prerequisite for the analysis of genetic variation for use in conservation and practical plant breeding. The development of genomic resources and tools for a crop is a crucial step both for the determination of genetic diversity and for the development of DNA markers associated with desirable traits for use in marker-aided breeding, and such resources and tools are often lacking for underutilized and minor crops. In the present study, useful and novel genetic information was gathered for noug despite the moderate success rate of the KASP genotyping.

4.1. The SNP/KASP Markers

SNP loci can be bi-, tri-, or tetra-allelic [60,61] but bi-allelic SNPs are the most commonly used because of their abundance and simplicity for use in genetic analyses. However, the level of polymorphism of bi-allelic SNPs may be lower, on average, when compared with other types of SNPs. Similar to other types of molecular markers, transcriptome-based SNPs are generally less polymorphic than non-genic SNPs. In this study, only 48.4% of the genotyped loci were polymorphic, which is not surprising as the SNPs are located in the exons of their corresponding genes and a limited number of accessions were analyzed. Polymorphism information content (PIC) and heterozygosity are common measures of polymorphism of a marker locus [62,63]. In the present study, the PIC of each SNP locus was calculated according to Hildebrand et al. [62]. In this approach, the maximum PIC value for bi-allelic SNPs is 0.375, and it is obtained when both alleles had a frequency of 0.5. The SNPs used in this study are bi-allelic, and their PIC values ranged from 0.01 to 0.37 with a mean of 0.24. Fifty percent of these SNP loci have a PIC value of more than 0.25 and hence are highly informative, and can be prioritized for various applications including for genetic diversity analysis of wider noug genetic resources and its weedy/wild close relatives. Two SNP-based studies in rice cultivars also reported a similar range in PIC values with a mean of 0.23 [64] and 0.28 [65].

In this study, 30% of the SNP loci have a minor allele frequency of less than 0.1, and as a result, the overall mean effective number of alleles was slightly below 1.5. The average observed (Ho) and expected (He) heterozygosities across the 202 loci were 0.25 and 0.29, respectively. In a study on a single noug population conducted using 43 EST-SSR markers, Dempewolf et al. [18] reported Ho and He values of 0.49 and 0.54, respectively. A separate study using a subset of these EST-SSR markers in 29 noug populations resulted in slightly lower values (Ho = 0.40 and He = 46) [22]. The average

Nei's gene diversity (Hs), an equivalent of expected heterozygosity, estimated based on random amplified polymorphic DNA (RAPD) markers [19] and amplified fragment length polymorphism (AFLP) markers [20] were 0.18 and 0.21, respectively. Similar levels of variation were obtained in *Guizotia scabra*, a closely related wild/weedy species [66]. These values are a bit lower than values obtained in the present study although the total number of alleles in the RAPD and AFLP based studies were 376 and 966, respectively [19,20]. This is partly because both RAPD and AFLP are dominant markers that generally underestimate the polymorphism level of a marker locus. The results reflect the advantage of co-dominant markers, such as SNPs and SSRs over dominant markers as have been reported in various previous publications [67–69]. On the other hand, the higher genetic variation reported in an EST-SSR (multi-allelic)-based study in noug [22] when compared to the present study (bi-allelic SNPs) is likely because of a higher effective number of alleles per locus (2.2) and a larger sample size of the former.

The fixation indices (F_{IT}, F_{IS}, and F_{ST}), also referred to as F-statistics, are measures of inbreeding in terms of total population (T), sub-populations (S), and individuals (I) for each locus [70,71]. All three indices have a maximum value of one. Negative values of F_{IS} and F_{IT} indicate excess heterozygosity and a value of one for these indices indicates 100% homozygosity in each sub-population. For F_{ST}, a measure of differentiation of sub-populations, the maximum value is attained when sub-populations are fixed for different alleles. In the present study, each accession is considered as a sub-population and the 24 accessions together form a total population. Noug is a strictly outcrossing species [11,44,72], and hence the overall observed heterozygosity (Ho) is expected to be higher than expected heterozygosity (He) if all other HWE assumptions are met. However, the mean Ho was less than the mean He for the polymorphic loci in the present study (Table 2; Table S3). Similarly, Ho was less than He, on average, in an EST-SSR-based study in noug [18,22]. Large variation was observed in fixation indices among the SNP loci. For example, F_{IS} values ranged from −0.90 to 1.00, indicating that some loci are in a state of heterozygote excess whereas some other loci lack heterozygotes. The mean values of F_{IT}, F_{IS}, and F_{ST} revealed in the present study were 0.13, 0.21, and 0.10 in that order. The values of these indices were 0.19, 0.15, and 0.04 in an EST-SSR based study [22]. The results suggest the overall low but significant heterozygote deficiency and low population differentiation in noug. The HWE test in this study revealed that about half of the loci (49.3%) showed significant deviation from HWE (Figure 2; Table S3). Interestingly, the vast majority of these loci (88%) showed heterozygote deficiency and only 12% showed heterozygote excess. The result clearly showed that different loci are under different kinds and levels of selection pressure, and other evolutionary forces, with overall heterozygote disadvantage in noug. Given that the study is based on genic SNPs, such a result is not unexpected.

Considering the 16 loci with a highly significant deviation from HWE (Table 1), 10 of the 12 loci that showed heterozygote excess lacked one of the two homozygous genotypes although their minor allele frequency (MAF) was as high as 0.47. This suggests that one of the two alleles in each locus makes the homozygous genotypes less fit, or the locus is in linkage disequilibrium (LD) with other locus or loci with a significant fitness value within its corresponding gene or other genes nearby. Among these loci, 3143A and 3143B are within the coding region of the 17.8 kDa class I heat shock protein-like gene. The SNP at the locus 3143A resulted in a Serine/Arginine non-synonymous substitution of the 99th amino acid of the protein coded by this gene. The 17.8 kDa class I heat shock protein is one of the small heat shock proteins in plants that are produced in response to high temperature stress [73]. Hence, further analysis may reveal variation in the response to heat stress among different genotypes of this locus. The SNPs that resulted in synonymous substitution are likely in LD with other locus or loci that affect the fitness of individual genotypes. Similarly, all four loci with heterozygote deficiency (Table 1) lacked heterozygous genotype despite having an MAF ranging from 0.19 to 0.47. However, all four resulted in synonymous amino acid substitution suggesting that these loci are tightly linked to a locus where strong heterozygote disadvantage is manifested. Detailed studies on the genes harboring these 16 SNPs using diverse noug germplasm may shed light on their functions.

4.2. Genetic Variation within Accessions

The 24 accessions used in the present study showed a relatively narrow range of variation in terms of percent polymorphic loci (PPL). The accessions with the lowest (NG108, 72%) and highest PPL (NG092, 85%) were from northeastern and eastern Ethiopia, respectively. Interestingly, both accessions are from areas less known for noug cultivation. In terms of genetic variation within accessions (I, He, and uHe), NG108 and NG106 are the lowest whereas NG092 and NG095 are the highest. These landrace populations are from different regions and hence the levels of genetic variation within landrace populations cannot be attributed to regions of cultivation. Similarly, NG092, NG095, and NG106 are landrace populations collected from middle-altitudes (Table S1), and the mean values of I, He, and uHe are similar for the three altitudinal groups, and hence altitude does not seem to have a significant effect on the extent of genetic variation. The estimates of genetic variation within accessions (I, He, and uHe) for the two cultivars (Fogera and Shambu), are close to the overall mean values of each parameter. Noug breeding in Ethiopia involves mass and recurrent selection [17]. Seemingly, the breeding methods used did not significantly affect the genetic variation of these cultivars. It could also be the case that the genetic variation in these cultivars has been increased (after their release) as a result of unintended gene flow from landraces grown in adjacent areas during seed multiplication or regeneration of the cultivars by the Ethiopian Institute of Agricultural Research at their field sites.

The 24 accessions showed several folds of variation in terms of fixation index (F) ranging from 0.013 (NG099) to 0.162 (NG092). Accession NG092 differs from most of the accessions not only in its higher genetic variation but also in having a larger deviation from HWE on average. On the other hand, the accessions are quite similar in terms of showing a low level of variation in diversity parameter Theta (H) that measures nucleotide diversity based on effective population size and mutation rate [52]. The similarly low Theta (H) values for the accessions are related to the case that they have a small variation in PPL, similar effective population size, and each accession has only two alleles at a polymorphic locus (bi-allelic SNPs).

Analysis of average evolutionary divergence over sequence pairs within populations using the Tajima-Nei model [53] produced interesting results. In order to evaluate the effect of allele frequencies in estimating evolutionary divergence between individuals within populations, this analysis was separately conducted for loci with MAF equal to or above 0.3 ($EAED_2$) and loci with MAF below 0.3 ($EAED_3$), in addition to their combined analysis ($EAED_1$). A slightly higher positive correlation ($r = 0.79$) for $EAED_1$ vs. $EAED_2$ than for $EAED_1$ vs. $EAED_3$ ($r = 0.75$) suggests that loci with MAF equal to or above 0.3 may have a higher effect on overall evolutionary divergence within populations than loci with a MAF below 0.3. The lack of correlation between $EAED_2$ and $EAED_3$ suggests that the selection of noug landrace populations for conservation and breeding should consider SNP loci with a MAF below 0.3 and equal to or above 0.3 separately, as different populations may be more valuable for conserving higher genetic diversity or in harboring desirable traits in the case of the two groups of loci. For example, accession NG111 is the highest in $EAED_2$ but the second-lowest in $EAED_3$, revealing that it has lower frequencies for MAF below 0.3 when compared to almost all other accessions. This is different from the case in accession NG103, where all three EAEDs are the lowest.

Noug accessions from low altitudes areas (1400 to 1680 masl) had the lowest $EAED_2$ (0.447) and highest $EAED_3$ (0.205) compared to the other two altitudinal areas. The results suggest that low-frequency alleles are more common in the lowland than in the highland areas of noug cultivation. This provides a slightly higher weight for lowland areas as sources of new alleles than higher altitude areas. The comparison of the two cultivars and the breeding population on one hand and the landrace populations on the hand clearly indicate that the breeding process did not have negative effects on the overall genetic divergence between individuals with populations, including for loci with a MAF below 0.3. This is probably because the loci included in the present study do not have an effect on the traits targeted for breeding and hence the alleles were not differentially selected.

4.3. Genetic Variation among Accessions and Population Structure

Partitioning the total genetic variation into the within and among populations components is an important step in understanding the genetic structure of populations and their adaptation to local environmental conditions. Outcrossing species generally tend to have higher genetic variation within populations than among populations, which is partly attributable to higher gene flow between populations than in the case of self-pollinating species [74,75]. Both the present study and previous DNA marker-based studies in noug [19,20] revealed higher genetic variation within populations than among populations. In the RAPD [19] and AFLP [20] based studies, about 35% and 23%, respectively, of the total variation accounted for population differentiation. In the present study, however, only 4.5% of the total variation differentiated the populations. The significantly lower population differentiation in the present study is mainly because it is based on transcriptome derived bi-allelic SNP markers that are generally more conserved than RAPD and AFLP-based markers.

In an EST-SSR-based study, Dempewolf et al. [22] reported a significant population differentiation accounting for 6% of the total variation, which is quite similar to the result of the present study. An EST-SSR based study in the Ethiopian potato (*Plectranthus edulis*), an outcrossing species, and a clonally propagating crop, also revealed very low genetic differentiation (2.5%) among populations [76]. Generally, outcrossing species have lower variation among populations than predominately self-pollinating species, as revealed using different marker systems in crops such as sorghum (83% [77,78]; 70% [79,80]), common beans (95% [81]), field peas (41% [82]), and durum wheat (31% [25]). However, it should be noted that self-pollinating species can have a low population differentiation when factors other than the mating system affecting the population structure are strong. For example, variation among populations accounted for only 13% and 6% of the total variation in arabica coffee [83] and korarima [84], respectively.

In agreement with previous noug research [19,20], there was no significant differentiation between the three altitudinal groups of accessions in the present study, as revealed by hierarchical AMOVA. However, alleles specific to each altitudinal group were present (Table S4). Four, three, and one specific allele(s) were recorded for low (1400 to 1680 masl), middle (1820 to 1968 masl) and high (2045 to 2590 masl) altitudinal groups. It is interesting to note the presence of more specific alleles in low-altitude areas than in high-altitude areas despite the fact that noug cultivation is more prominent in high-altitude areas. The result suggests the importance of including low-altitude populations in the noug breeding program. Similarly, there was no significant differentiation between the six regional groups of accessions although the grouping was made under the assumption that higher access to gene flow exists within the groups than among the groups. The marginally significant differentiation between accessions from the Oromia region and Amhara-Tigray region may suggest a slightly higher level of germplasm exchange within the regions than the nation-wide average.

The cluster analysis at the individual genotype level produced three major clusters (Figure 3). However, there was no clear clustering pattern of genotypes according to their accession, which is in line with the low population differentiation obtained through AMOVA. The lack of significant differentiation among altitudinal and regional groups revealed through AMOVA was also evident in the cluster analysis at the level of accessions (Figure 4). This is the case even when only loci with a MAF below 0.3 (loci close to fixation) was used, suggesting a wide distribution of low-frequency alleles. In the PCoA that explained 44% of the total variation in its first two principal axes, the accessions were tightly clustered together with the exception of NG096, NG097, and NG124. Accession NG097 has the highest mean pair-wise F_{ST} (0.076) followed by NG096 (0.074; Table 4). Accession NG097 lacked alleles at two loci that exist in all other landrace accessions (loci 3926 and 12963A). It also lacked a number of alleles present in the vast majority of accessions. Similarly, NG096 lacked alleles common to most accessions at a number of loci. Unlike other landrace accessions, these two accessions were collected from areas where noug cultivation is rare, and their higher genetic differentiation could be the result of isolation by distance [85] by being kept at a household/local community-level for a long time. Accession NG124 is a breeding population developed through crossbreeding of selected

genotypes from different landrace populations for improved oil and oleic acid content. Unlike all other accessions, it lacks an allele at a locus (locus 1117) within a gene that codes for tobamovirus multiplication protein 1-like protein. Tobamovirus multiplication protein 1 is required for efficient multiplication of a tobamovirus in *Arabidopsis* [86,87]. Further research on this locus is necessary to determine its function in noug.

Various approaches have been used to determine the source of individual genotypes in population genetics research [54,88–90]. Pritchard et al. [54] developed a model-based method of population structure analysis for multi-locus genotypic data, in which populations that are characterized by a set of allele frequencies across loci are assumed. In this approach, individual members of predefined populations are probabilistically assigned to a single cluster (inferred population) or receive joint assignment to more than one cluster if the model determines that they are admixed. In this study, we used 24 predefined populations (accessions) to determine the population genetic structure in noug using this model. The analysis using the ΔK method of Evanno et al. [56] showed that the individuals within these predefined populations most likely originated from three genetic populations (K = 3). Interestingly, all individuals across the predefined populations are the results of admixture from the three genetic populations although it is to a different extent (ranging from 13 to 55% on average, data not shown). This and other points discussed above generally show weak population structure in noug due to population admixture caused by strong gene flow between populations via pollen and a step by step nation-wide germplasm exchange.

5. Conclusions

In the present study, transcriptome sequencing was conducted, and novel SNP markers were identified in noug. More than 50% of these SNPs were successfully converted to KASP markers and used for the genotyping of noug populations collected from wide geographic areas in Ethiopia. Polymorphic SNP/KASP markers were used for population genetics analyses. The loci with high PIC or showing highly significant deviation from HWE should be prioritized for further research in this crop. The study revealed low but significant differentiation between noug populations. Nonetheless, there was generally a lack of, or poor differentiation between noug at different altitudinal ranges or regions level. This indicates strong gene flow between populations grown at different altitudes and regions, particularly between major noug cultivation areas. However, this study also gave clear indications that landrace populations cultivated in isolated areas deviate in genetic diversity from those populations in common noug cultivation areas. Thus, such populations from isolated areas in Ethiopia may be sources of useful mutations, and hence should be considered for their conservation and use in breeding of this oil crop. Overall, the transcript sequences and the SNP and KASP markers developed in this study are highly useful resources for applications such as population genetics analyses and genome-wide association research. However, the markers are relatively small in number and were developed based on only two genotypes. Hence, methods such as RNAseq based on a larger number of diverse germplasm should be applied to develop DNA markers, in the thousands, that can have wide applications by avoiding potential drawbacks associated with ascertainment bias. This approach facilitates the identification of markers associated with traits of interest through genetic linkage and association mapping, for their use in marker-aided breeding.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/11/11/1373/s1, Table S1: Accession code and type, and altitude, region, and location of collecting sites, Table S2: GenBank accession numbers of the 628 ESTs containing 959 SNPs initially used in the development of the KASP markers, and grouping of the SNP loci according to the results of the KASP analysis, Table S3: Polymorphic information content (PIC), gene diversity (H), fixation indices, gene flow (Nm), observed (Ho) and expected (He) heterozygosity, and *p*-value for deviation of each locus from HWE, for the 202 polymorphic SNP loci, Table S4: Population size (N) and frequency of alleles (A, C, G, or T) at the 202 polymorphic loci for the three altitudinal groups (Alt-1, 2, 3) formed from the 21 landrace accessions.

Author Contributions: Conceptualization, M.G., R.O., and S.T.; methodology, M.G., R.O., and S.T.; data analysis, S.T., M.G., and C.H.; writing—original draft preparation, S.T.; writing—review and editing, M.G., R.O., E.J., E.B.,

K.T., C.H., and S.T.; funding acquisition, M.G., E.B., K.T., E.J., and R.O.; supervision, R.O., M.G., E.J., E.B., and K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work is part of the development research project 2014-03517 financed by the Swedish Research Council (Vetenskapsrådet, VR), to which the authors are grateful.

Acknowledgments: We would like to thank Loren H. Rieseberg for the fruitful collaboration that led to the transcriptome sequencing of noug genotypes and the development of SNPs within the Compositae Genome Project.

Conflicts of Interest: The authors declare no conflict of interest and the funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Dagne, K.; Heneen, W.K. The karyotype and nucleoli of *Guizotia abyssinica* (Compositae). *Hereditas* **2008**, *117*, 73–83. [CrossRef]
2. Hiremath, S.C.; Murthy, H.N. Cytological studies in *Guizotia* (Asteraceae). *Caryologia* **1992**, *45*, 69–82. [CrossRef]
3. Dagne, K. Karyotypes, C-banding and nucleolar numbers in *Guizotia* (Compositae). *Plant Syst. Evol.* **1995**, *195*, 121–135. [CrossRef]
4. Hiremath, S.C.; Murthy, H.N.; Salimath, S.S. Quantitative nuclear DNA differences associated with genome evolution in *Guizotia* (Compositae). *Genetica* **1992**, *85*, 241–247. [CrossRef]
5. Ohta, T. Further Examples of Evolution by Gene Duplication Revealed through DNA Sequence Comparisons. *Genetics* **1994**, *138*, 1331–1337.
6. Meyerowitz, E.M. The first completely sequenced plant chromosomes, from the mustard *Arabidopsis thaliana*, reveal a dynamic genome that is constantly being rearranged. *Nature* **1999**, *402*, 731–732. [CrossRef]
7. Bennetzen, J.L.; Ma, J.; Devos, K.M. Mechanisms of Recent Genome Size Variation in Flowering Plants. *Ann. Bot.* **2005**, *95*, 127–132. [CrossRef]
8. Weiss, E.A. *Oil Seed Crops*; Longman Inc.: New York, NY, USA, 1983.
9. Murthy, H.N.; Hiremath, S.C.; Salimath, S.S. Origin, evolution and genome differentiation in *Guizotia abyssinica* and its wild species. *Theor. Appl. Genet.* **1993**, *87*, 587–592. [CrossRef]
10. Getinet, A.; Sharma, S.M. Niger (*Guizotia abyssinica* (L.F.)) Cass. In *Promoting the Conservation and Use of Underutilized and Neglected Crops*; International Plant Genetic Resources Institute (IPGRI): Rome, Italy, 1996.
11. Geleta, M.; Ortiz, R. The importance of *Guizotia abyssinica* (niger) for sustainable food security in Ethiopia. *Genet. Resour. Crop. Evol.* **2013**, *60*, 1763–1770. [CrossRef]
12. Geleta, M.; Stymne, S.; Bryngelsson, T. Variation and inheritance of oil content and fatty acid composition in niger (*Guizotia abyssinica*). *J. Food Compos. Anal.* **2011**, *24*, 995–1003. [CrossRef]
13. Dagne, K.; Jonsson, A. Oil Content and Fatty Acid Composition of Seeds of *Guizotia* Cass (Compositae). *J. Sci. Food Agric.* **1997**, *73*, 274–278. [CrossRef]
14. Ramadan, M.F.; Ramadan, M.F. Functional Properties, Nutritional Value, and Industrial Applications of Niger Oilseeds (*Guizotia abyssinica* Cass.). *Crit. Rev. Food Sci. Nutr.* **2012**, *52*, 1–8. [CrossRef] [PubMed]
15. Deme, T.; Haki, G.D.; Retta, N.; Woldegiorgis, A.Z.; Geleta, M. Mineral and Anti-Nutritional Contents of Niger Seed (*Guizotia abyssinica* (L.f.) Cass., Linseed (*Linum usitatissimum* L.) and Sesame (*Sesamum indicum* L.) Varieties Grown in Ethiopia. *Foods* **2017**, *6*, 27. [CrossRef] [PubMed]
16. Geleta, M.; Asfaw, Z.; Bekele, E.; Teshome, A. Edible oil crops and their integration with the major cereals in North Shewa and South Welo, Central Highlands of Ethiopia: An ethnobotanical perspective. *Hereditas* **2002**, *137*, 29–40. [CrossRef] [PubMed]
17. Alemaw, G.; Alamayehu, N. Highland oilcrops: A two-decade research experience in Ethiopia. In *Research Report No. 30*; Institute of Agricultural Research: Addis Ababa, Ethiopia, 1997.
18. Dempewolf, H.; Kane, N.C.; Ostevik, K.L.; Geleta, M.; Barker, M.S.; Lai, Z.; Stewart, M.L.; Bekele, E.; Engels, J.M.M.; Cronk, Q.C.B.; et al. Establishing genomic tools and resources for *Guizotia abyssinica* (L.f.) Cass.-the development of a library of expressed sequence tags, microsatellite loci, and the sequencing of its chloroplast genome. *Mol. Ecol. Resour.* **2010**, *10*, 1048–1058. [CrossRef] [PubMed]

19. Geleta, M.; Bryngelsson, T.; Bekele, E.; Dagne, K. Genetic diversity of *Guizotia abyssinica* (L.f.) Cass. (Asteraceae) from Ethiopia as revealed by random amplified polymorphic DNA (RAPD). *Genet. Resour. Crop. Evol.* **2006**, *54*, 601–614. [CrossRef]
20. Geleta, M.; Bryngelsson, T.; Bekele, E.; Dagne, K. Assessment of genetic diversity of *Guizotia abyssinica* (L.f.) Cass. (Asteraceae) from Ethiopia using amplified fragment length polymorphism. *Plant Genet. Resour.* **2008**, *6*, 41–51. [CrossRef]
21. Petros, Y.; Merker, A.; Zeleke, H. Analysis of genetic diversity of *Guizotia abyssinica* from Ethiopia using inter simple sequence repeat markers. *Hereditas* **2007**, *144*, 18–24. [CrossRef]
22. Dempewolf, H.; Tesfaye, M.; Teshome, A.; Bjorkman, A.D.; Andrew, R.L.; Scascitelli, M.; Black, S.; Bekele, E.; Engels, J.M.M.; Cronk, Q.C.B.; et al. Patterns of domestication in the Ethiopian oil-seed crop noug (*Guizotia abyssinica*). *Evol. Appl.* **2015**, *8*, 464–475. [CrossRef]
23. Ertiro, B.T.; Ogugo, V.; Worku, M.; Das, B.; Olsen, M.; Labuschagne, M.; Semagn, K. Comparison of Kompetitive Allele Specific PCR (KASP) and genotyping by sequencing (GBS) for quality control analysis in maize. *BMC Genom.* **2015**, *16*, 1–12. [CrossRef]
24. Shi, Z.; Liu, S.; Noe, J.; Arelli, P.; Meksem, K.; Li, Z. SNP identification and marker assay development for high-throughput selection of soybean cyst nematode resistance. *BMC Genom.* **2015**, *16*, 1–12. [CrossRef]
25. Kabbaj, H.; Sall, A.T.; Al-Abdallat, A.; Geleta, M.; Amri, A.; Filali-Maltouf, A.; Belkadi, B.; Ortiz, R.; Bassi, F.M. Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange. *Front. Plant Sci.* **2017**, *8*, 1277. [CrossRef]
26. Contreras-Soto, R.I.; Mora, F.; De Oliveira, M.A.R.; Higashi, W.; Scapim, C.A.; Schuster, I. A Genome-Wide Association Study for Agronomic Traits in Soybean Using SNP Markers and SNP-Based Haplotype Analysis. *PLoS ONE* **2017**, *12*, e0171105. [CrossRef] [PubMed]
27. Qu, C.; Jia, L.; Fu, F.; Zhao, H.; Lu, K.; Wei, L.; Xu, X.; Liang, Y.; Li, S.; Wang, R.; et al. Genome-wide association mapping and Identification of candidate genes for fatty acid composition in *Brassica napus* L. using SNP markers. *BMC Genom.* **2017**, *18*, 232. [CrossRef]
28. Liu, R.; Gong, J.; Xiao, X.; Zhang, Z.; Li, J.; Liu, A.; Lu, Q.; Shang, H.; Shi, Y.; Ge, Q.; et al. GWAS Analysis and QTL Identification of Fiber Quality Traits and Yield Components in Upland Cotton Using Enriched High-Density SNP Markers. *Front. Plant Sci.* **2018**, *9*, 1067. [CrossRef] [PubMed]
29. Belaj, A.; De La Rosa, R.; Lorite, I.J.; Mariotti, R.; Cultrera, N.G.; Beuzón, C.R.; González-Plaza, J.J.; Muñoz-Merida, A.; Trelles, O.; Baldoni, L. Usefulness of a New Large Set of High Throughput EST-SNP Markers as a Tool for Olive Germplasm Collection Management. *Front. Plant Sci.* **2018**, *9*, 1320. [CrossRef]
30. Kim, B.; Hwang, I.S.; Lee, H.J.; Lee, J.M.; Seo, E.; Choi, D.; Oh, C.-S. Identification of a molecular marker tightly linked to bacterial wilt resistance in tomato by genome-wide SNP analysis. *Theor. Appl. Genet.* **2018**, *131*, 1017–1030. [CrossRef]
31. Geleta, M.; Gustafsson, C.; Glaubitz, J.C.; Ortiz, R. High-Density Genetic Linkage Mapping of *Lepidium* Based on Genotyping-by-Sequencing SNPs and Segregating Contig Tag Haplotypes. *Front. Plant Sci.* **2020**, *11*, 448. [CrossRef]
32. Palmé, A.E.; Hagenblad, J.; Solberg, S.Ø.; Aloisi, K.; Artemyeva, A.M. SNP Markers and Evaluation of Duplicate Holdings of *Brassica oleracea* in Two European Genebanks. *Plants* **2020**, *9*, 925. [CrossRef]
33. Zanotto, S.; Vandenberg, A.; Khazaei, H. Development and validation of a robust KASP marker for zt2 locus in faba bean (*Vicia faba*). *Plant Breed.* **2019**, *139*, 375–380. [CrossRef]
34. Islam, A.S.M.F.; Blair, M.W. Molecular Characterization of Mung Bean Germplasm from the USDA Core Collection Using Newly Developed KASP-based SNP Markers. *Crop. Sci.* **2018**, *58*, 1659–1670. [CrossRef]
35. Grimm, K.D.S.; Porter, L.D. Development and Validation of KASP Markers for the Identification of Pea seedborne mosaic virus Pathotype P1 Resistance in *Pisum sativum*. *Plant Dis.* **2020**, *104*, 1824–1830. [CrossRef] [PubMed]
36. Leal-Bertioli, S.C.M.; Cavalcante, U.; Gouveia, E.G.; Ballén-Taborda, C.; Shirasawa, K.; Guimarães, P.M.; Jackson, S.A.; Bertioli, D.J.; Moretzsohn, M.C. Identification of QTLs for Rust Resistance in the Peanut Wild Species *Arachis magna* and the Development of KASP Markers for Marker-Assisted Selection. *G3 Genes. Genom. Genet.* **2015**, *5*, 1403–1413. [CrossRef]
37. Kang, J.-W.; Lee, S.-B.; Lee, J.-Y.; Kwon, Y.-H.; Lee, S.-M.; Rolly, N.K.; Shin, D.; Cha, J.-G.; Park, D.-S.; Ko, J.-M.; et al. Development and Validation of KASP Markers for *Stv-bi*, a Rice Stripe Virus Resistance Gene in Rice (*Oryza sativa* L.). *Plant Breed. Biotechnol.* **2020**, *8*, 196–201. [CrossRef]

38. Cheon, K.-S.; Baek, J.; Cho, Y.-I.; Jeong, Y.-M.; Lee, Y.-Y.; Oh, J.; Won, Y.J.; Kang, D.-Y.; Oh, H.; Kim, S.L.; et al. Single Nucleotide Polymorphism (SNP) Discovery and Kompetitive Allele-Specific PCR (KASP) Marker Development with Korean Japonica Rice Varieties. *Plant Breed. Biotechnol.* **2018**, *6*, 391–403. [CrossRef]
39. Han, G.; Liu, S.; Jin, Y.; Jia, M.; Ma, P.; Liu, H.; Wang, J.; An, D.-G. Scale development and utilization of universal PCR-based and high-throughput KASP markers specific for chromosome arms of rye (*Secale cereale* L.). *BMC Genom.* **2020**, *21*, 1–14. [CrossRef]
40. Burow, G.; Chopra, R.; Sattler, S.; Burke, J.; Acosta-Martinez, V.; Xin, Z. Deployment of SNP (CAPS and KASP) markers for allelic discrimination and easy access to functional variants for brown midrib genes *bmr6* and *bmr12* in *Sorghum bicolor*. *Mol. Breed.* **2019**, *39*, 115. [CrossRef]
41. Wu, J.; Wang, Q.; Kang, Z.; Liu, S.; Li, H.; Mu, J.; Dai, M.; Han, D.; Zeng, Q.; Chen, X. Development and Validation of KASP-SNP Markers for QTL Underlying Resistance to Stripe Rust in Common Wheat Cultivar P10057. *Plant Dis.* **2017**, *101*, 2079–2087. [CrossRef]
42. Fang, T.; Lei, L.; Li, G.; Powers, C.; Hunger, R.M.; Carver, B.F.; Yan, L. Development and deployment of KASP markers for multiple alleles of Lr34 in wheat. *Theor. Appl. Genet.* **2020**, *133*, 2183–2195. [CrossRef]
43. Makhoul, M.; Rambla, C.; Voss-Fels, K.P.; Hickey, L.T.; Snowdon, R.J.; Obermeier, C. Overcoming polyploidy pitfalls: A user guide for effective SNP conversion into KASP markers in wheat. *Theor. Appl. Genet.* **2020**, *133*, 2413–2430. [CrossRef]
44. Geleta, M.; Bryngelsson, T. Population genetics of self-incompatibility and developing self-compatible genotypes in *Guizotia abyssinica* (L.f.) Cass. (Asteraceae). *Euphytica* **2010**, *176*, 417–430. [CrossRef]
45. Hodgins, K.A.; Lai, Z.; De Oliveira, L.O.; Still, D.W.; Scascitelli, M.; Barker, M.S.; Kane, N.C.; Dempewolf, H.; Kozik, A.; Kesseli, R.V.; et al. Genomics of Compositae crops: Reference transcriptome assemblies and evidence of hybridization with wild relatives. *Mol. Ecol. Resour.* **2013**, *14*, 166–177. [CrossRef] [PubMed]
46. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]
47. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Subgroup, G.P.D.P. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
48. Yeh, F.; Yang, R.; Boyle, T. *POPGENE: Microsoft Window-Based Freeware for Population Genetic Analysis*; Version 1.31; University of Alberta: Edmonton, AB, Canada, 1999.
49. Peakall, R.; Smouse, P.E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research–an update. *Bioinformatics* **2012**, *28*, 2537–2539. [CrossRef]
50. Excoffier, L.; Lischer, H.E.L. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **2010**, *10*, 564–567. [CrossRef]
51. Kumar, S.; Stecher, G.; Tamura, K.K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef]
52. Nei, M. *Molecular Evolutionary Genetics*; Columbia University Press: New York, NY, USA, 1987.
53. Tajima, F.; Nei, M. Estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **1984**, *1*, 269–285. [CrossRef]
54. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **2000**, *155*, 945–959.
55. Li, Y.-L.; Liu, J.-X. StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol. Ecol. Resour.* **2017**, *18*, 176–177. [CrossRef]
56. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]
57. Kopelman, N.M.; Mayzel, J.; Jakobsson, M.; Rosenberg, N.A.; Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **2015**, *15*, 1179–1191. [CrossRef] [PubMed]
58. Moragues, M.; Comadran, J.; Waugh, R.; Milne, I.; Flavell, A.J.; Russell, J.R. Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theor. Appl. Genet.* **2010**, *120*, 1525–1534. [CrossRef] [PubMed]
59. Nielsen, R. Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* **2000**, *154*, 931–942. [PubMed]

60. Chang, H.-W.; Cheng, Y.-H.; Chuang, L.-Y.; Yang, C.-H. SNP-RFLPing 2: An updated and integrated PCR-RFLP tool for SNP genotyping. *BMC Bioinform.* **2010**, *11*, 173. Available online: http://www.biomedcentral.com/1471-2105/11/173 (accessed on 19 November 2020). [CrossRef]
61. Evans, J.; Kim, J.; Childs, K.L.; Vaillancourt, B.; Crisovan, E.; Nandety, A.; Gerhardt, D.J.; Richmond, T.A.; Jeddeloh, J.A.; Kaeppler, S.M.; et al. Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass *Panicum virgatum*. *Plant J.* **2014**, *79*, 993–1008. [CrossRef]
62. Hildebrand, C.E.; Torney, D.V.; Wagner, R.P. Informativeness of polymorphic DNA markers. In *The Human Genome Project: Decipering the Blueprint of Heredity (Cooper NG edited)*; University Science Books: Mill Valley, CA, USA, 1994; pp. 100–102.
63. Shete, S.; Tiwari, H.; Elston, R.C. On Estimating the Heterozygosity and Polymorphism Information Content Value. *Theor. Popul. Biol.* **2000**, *57*, 265–271. [CrossRef]
64. Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.K.; Singh, R. Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. *PLoS ONE* **2013**, *8*, e84136. [CrossRef]
65. Chen, H.; He, H.; Zou, Y.; Chen, W.; Yu, R.; Liu, X.; Yang, Y.; Gao, Y.-M.; Xu, J.-L.; Fan, L.-M.; et al. Development and application of a set of breeder-friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **2011**, *123*, 869–879. [CrossRef]
66. Geleta, M.; Bryngelsson, T.; Bekele, E.; Dagne, K. AFLP and RAPD analyses of genetic diversity of wild and/or weedy *Guizotia* (Asteraceae) from Ethiopia. *Hereditas* **2007**, *144*, 53–62. [CrossRef]
67. Milligan, B.G.; McMurry, C.K. Dominant vs. codominant genetic markers in the estimation of male mating success. *Mol. Ecol.* **1993**, *2*, 275–283. [CrossRef]
68. Bonnett, D.; Rebetzke, G.; Spielmeyer, W. Strategies for efficient implementation of molecular markers in wheat breeding. *Mol. Breed.* **2005**, *15*, 75–85. [CrossRef]
69. Foll, M.; Gaggiotti, O.E. A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* **2008**, *180*, 977–993. [CrossRef]
70. Wright, S. The genetical structure of populations. *Ann. Eugen.* **1949**, *15*, 323–354. [CrossRef]
71. Wright, S. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* **1965**, *19*, 395. Available online: http://www.jstor.com/stable/2406450 (accessed on 19 November 2020). [CrossRef]
72. Nemomissa, S.; Bekele, E.; Dagne, K. Self-incompatibility system in the Ethiopian populations of *Guizotia abyssinica* (L.F.) Cass. (niger). *SINET Ethiop. J. Sci.* **1999**, *22*, 67–88. [CrossRef]
73. Waters, E.R. The Molecular Evolution of the Small Heat-Shock Proteins in Plants. *Genetics* **1995**, *141*, 785–795.
74. Hamrick, J.L.; Godt, M.J.W. Effects of life history traits on genetic diversity in plant species. *Philos. Trans. R. Soc. B Biol. Sci.* **1996**, *351*, 1291–1298. [CrossRef]
75. Nybom, H. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol. Ecol.* **2004**, *13*, 1143–1155. [CrossRef]
76. Gadissa, F.; Tesfaye, K.; Dagne, K.; Geleta, M. Genetic diversity and population structure analyses of *Plectranthus edulis* (Vatke) Agnew collections from diverse agro-ecologies in Ethiopia using newly developed EST-SSRs marker system. *BMC Genet.* **2018**, *19*, 92. [CrossRef]
77. Ng'uni, D.; Geleta, M.; Hofvander, P.; Fatih, M.; Bryngelsson, T. Comparative genetic diversity and nutritional quality variation among some important Southern African sorghum accessions [*Sorghum bicolor* (L.) Moench]. *Aust. J. Crop. Sci.* **2012**, *6*, 56–64.
78. Ng'Uni, D.; Geleta, M.; Bryngelsson, T. Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). *Hereditas* **2011**, *148*, 52–62. [CrossRef] [PubMed]
79. Motlhaodi, T.; Geleta, M.; Bryngelsson, T.; Fatih, M.; Chite, S.; Ortiz, R. Genetic diversity in ex-situ conserved sorghum accessions of Botswana as estimated by microsatellite markers. *Aust. J. Crop. Sci.* **2014**, *8*, 35–43.
80. Motlhaodi, T.; Geleta, M.; Chite, S.; Fatih, M.; Ortiz, R.; Bryngelsson, T. Genetic diversity in sorghum germplasm from Southern Africa as revealed by microsatellite markers and agromorphological traits. *Genet. Resour. Crop. Evol.* **2017**, *64*, 599–610. [CrossRef]
81. Hegay, S.; Geleta, M.; Bryngelsson, T.; Gustavsson, L.; Hovmalm, H.P.; Ortiz, R. Comparing genetic diversity and population structure of common beans grown in Kyrgyzstan using microsatellites. *Sci. J. Crop. Sci.* **2012**, *1*, 63–75.

82. Teshome, A.; Bryngelsson, T.; Dagne, K.; Geleta, M. Assessment of genetic diversity in Ethiopian field pea (*Pisum sativum* L.) accessions with newly developed EST-SSR markers. *BMC Genet.* **2015**, *16*, 1–12. [CrossRef]
83. Geleta, M.; Herrera, I.; Monzón, A.; Bryngelsson, T. Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers. *Sci. World J.* **2012**, *2012*, 1–11. [CrossRef]
84. Chombe, D.; Bekele, E.; Bryngelsson, T.; Teshome, A.; Geleta, M. Genetic structure and relationships within and between cultivated and wild korarima [*Aframomum corrorima* (Braun) P.C.M. Jansen] in Ethiopia as revealed by simple sequence repeat (SSR) markers. *BMC Genet.* **2017**, *18*, 1–13. [CrossRef]
85. Wright, S. Isolation by Distance. *Genetics* **1943**, *28*, 114–138.
86. Ishikawa, M.; Naito, S.; Ohno, T. Effects of the *tom1* mutation of *Arabidopsis thaliana* on the multiplication of tobacco mosaic virus RNA in protoplasts. *J. Virol.* **1993**, *67*, 5328–5338. [CrossRef]
87. Yamanaka, T.; Ohta, T.; Takahashi, M.; Meshi, T.; Schmidt, R.; Dean, C.; Naito, S.; Ishikawa, M. *TOM1*, an *Arabidopsis* gene required for efficient multiplication of a tobamovirus, encodes a putative transmembrane protein. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10107–10112. [CrossRef] [PubMed]
88. Rannala, B.; Mountain, J.; Rannala, B.; Mountain, J.L. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 9197–9201. [CrossRef] [PubMed]
89. Davies, N.; Villablanca, F.X.; Roderick, G.K. Determining the source of individuals: Multilocus genotyping in nonequilibrium population genetics. *Tree* **1999**, *14*, 17–21. [CrossRef]
90. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [CrossRef]

Acta Universitatis agriculturae Sueciae
Doctoral Thesis No. 2023:100

The doctoral thesis extensively investigated noug (*Guizotia abyssinica*) nutritional profiling, the development of cutting-edge genomic tools for application in population genetic studies, and the conversion of Single Nucleotide Polymorphisms (SNPs) into molecular markers referred to as competitive allele-specific PCR (KASP). The genomic tool development process involved the validation of robust markers that significantly contributed to population genetic studies. Furthermore, the research delved into examining genotype-by-environment interactions. The findings are applicable in population genetic studies and diversity analysis, genetic resource conservation, pre-breeding, and breeding efforts.

Sewalem Tsehay Wondim received his Doctor of Philosophy in Agricultural Sciences (Plant Breeding and Genetics) from the Department of Plant Breeding, Swedish University of Agricultural Sciences (SLU), Alnarp, Sweden. He received his MSc in Applied Genetics and BSc in Applied Biology from Addis Ababa University, Addis Ababa, Ethiopia.

Acta Universitatis agriculturae Sueciae presents doctoral theses from the Swedish University of Agricultural Sciences (SLU).

SLU generates knowledge for the sustainable use of biological natural resources. Research, education, extension, as well as environmental monitoring and assessment are used to achieve this goal.

ISSN 1652-6880
ISBN (print version) 978-91-8046-252-5
ISBN (electronic version) 978-91-8046-253-2