MilkOligoCorpus annotation guidelines

Mathilde Rumeau, Marine Courtin, Robert Bossy, Clara Sauvion, Valentin Loux, Mouhamadou Ba, Christelle Knudsen, Sylvie Combes, Claire Nédellec, Louise Deleger

MilkOligoCorpus annotation guidelines

HoloOLIGO project ANR-21-CE20-0045

Mathilde Rumeau[1], Marine Courtin[2], Robert Bossy[2], Clara Sauvion[2], Valentin Loux[2,3] Mouhamadou Ba[2,3], Christelle Knudsen[1], Sylvie Combes[1], Claire Nédellec[2] and Louise Deléger[2]

[1] GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France

[2] Université Paris-Saclay, INRAE, MaIAGE, Jouy-en-Josas, France

[3] Université Paris-Saclay, INRAE, BioinfOmics, MIGALE Bioinformatics Facility, Jouy-en-Josas, France

December 4, 2024

Table of contents

1. Introduction

a. Purpose

This document describes the guidelines for annotating the MilkOligoCorpus. The goal is to design a corpus to be used for evaluating and training extraction methods of information related to milk oligosaccharides (MO) of different mammal species from the literature. Targeted information include: MO (with quantification when available), species mentions (along with the breed and the number of individuals from the experimentation when available), stage of lactation, maternal metadata, methodology and type of samples used in the study and geography, if available.

b. Conventions

A sample of articles has been annotated in order to complete these guidelines. The entity and binary relationship annotations along with the annotations of n-ary relationships (more than two arguments) are available at https://doi.org/10.57745/LFXGFO. Examples extracted from these articles are given in *italic font*, followed by explanations. Each example is given with its reference name and annotated with the same legend as on this annotation editor. The following legend describes how entities are highlighted:

Species
Breed
Geography
Female physiological stage
Individual number
Sample type
Lactation stage
Postpartum age
Methodology of analysis
Oligosaccharide's type
Oligosaccharide's name
Absolute quantification
Relative quantification
Oligosaccharide's richness

When words are highlighted with two different colors, it means they belong to two separate entity types.
Each entity type has its own description section. This description includes a definition of the entity type, the ontology used for the entity mention normalization (if the entity must be normalized) and annotation rules for human annotators. These sections are presented sequentially, although examples typically include references to multiple entity types.
Following entity type descriptions, a dedicated section details the relations between them. These relations can be binary or involve more than two entities.

2. Entities referring to individuals

a. Species

i. Definition

Species belonging to the mammal group are the vertebrate animals in which the young are nourished with milk from special mammary glands of the mother. In this study we have decided to exclude marsupials (kangaroos, opossums and wallabies). Species names are normalized by relevant sub-tree taxa from the NCBI Taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy). The species identifier assigned to each occurrence of a species entity is taken from the NCBI Taxonomy (For instance with pig, NCBI ID: 9823).

ii. Synonyms

Several different words can be used to designate the same species. They must all be annotated.

A total of 488 breast milk samples were collected from 335 healthy mothers at five different time points (Liu_2021_part-1) (Liu et al., 2021)

In this example, “mothers” is used to designate humans.

iii. Repetition

Frequently, the species name mentioned at the beginning of the document differs from subsequent references to the species. All species references must be annotated and assigned to the same identifier in the taxonomy.

Previously, 19 fucosylated PMOs were identified in porcine milk ... due to different breeds of swine, stage of milk collection, parity of swine milk and methodology used for MO analysis. (Wei_2018_part-2) (Wei et al., 2018)

Porcine and swine refer to the same species, they are both annotated

Bovine colostrum from Holstein-Friesian cows and porcine colostrum from Landrace pigs were obtained on-site at Teagasc Food Research Centre, Moorepark (Fermoy, Cork, Ireland). (Albrecht_2014_part-2) (Albrecht et al., 2014)

Bovine and cows both refer to *Bos Taurus* species while porcine and pigs designate the *Sus Scrofa* one.

Milk samples from 7 individual dogs were used in the present study. To prepare for milk collection, lactating bitches were separated from their suckling pups ... (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Another example is the dog. In the same text we can find general and female mentions.

iv. Succession of species names

When several terms designating the same species follow each other, they must be annotated separately.

The only non-human primates whose milk sugars have been studied appear to be the rhesus monkey and the brown capuchin (Cebus apella). (Urashima_2001_part-1) (Urashima et al., 2001)

The common species name is followed by the scientific name. They both refer to the same animal and must be annotated.

v. Order, Sub-order, Family

Those 3 taxonomic ranks are too general, including multiple species, thus they must not be annotated.

Several OS were previously identified from milk of different species from the order of the Carnivora. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, the word order is clearly stated.

Structural characterization of neutral and acidic oligosaccharides in the milks of strepsirrhine primates: greater galago, aye-aye, Coquerel's sifaka and mongoose lemur (Taufik_2012_part-1) (Taufik et al., 2012)

Strepsirrhine primates is a suborder, therefore it must not be annotated.

Neutral and acidic oligosaccharides were isolated from milk of the greater galago (Galagidae: Otolemur crassicaudatus), aye-aye (Daubentoniidae: Daubentonia madagascariensis), Coquerel's sifaka (Indriidae: Propithecus coquereli) and mongoose lemur (Lemuridae: Eulemur mongoz), ... (Taufik_2012_part-1) (Taufik et al., 2012)

Before the scientific name of each species, family names are mentioned, they must not be annotated.

vi. Subspecies

On the contrary, subspecies belong to a lower rank. Thus, when they are mentioned, they must be annotated. They are ranked as subspecies in the NCBI Taxonomy.

Isoglobotriose, which had previously been found to be a dominant saccharide in mature milk from the Ezo brown bear, the Japanese black bear and the polar bear, was not found in the polar bear colostrum. (Urashima_2003a_part-1) (Urashima et al., 2003a)

Japanese black bear is a subspecies of Asiatic black bear in NCBI Taxonomy, it must be annotated. Even though Ezo brown bear is not registered in the NCBI taxonomy, we must annotate it. We will assign it with the most accurate taxon available in NCBI.

vii. General terms

General terms that do not refer to specific mammal species should not be annotated.

Milk from domestic animals contained a much larger variety of complex oligosaccharides (Albrecht_2014_part-2) (Albrecht et al., 2014)

Domestic animals include multiple species; therefore, it is impossible to know precisely which species are concerned.

viii. Young animals

As we are looking for data about mother's milk, we are only interested in animals that are mothers. However, sometimes articles are studying milk effects on newborns. Thus, the newborn designation must not be annotated. Nevertheless, several terms may be encountered in the articles:

- The young animal may be called by a specific name designed especially for young animals (lamb, piglet, calf...): they must not be annotated.

Individual fecal samples were collected from three piglets, 1–2 d and 1 w old (Proefaccommodatie de Tolakker), within 24 h after porcine milk ingestion. Fecal samples were stored at −80 °C until use. An overview of matching porcine colostrum, mature milk samples, and piglet fecal samples is displayed in Table 1. (Difilippo_2016_part-2) (Difilippo et al., 2016)

In this example, we do not annotate piglet, which correspond to newborn pig but we do annotate porcine since it designates the pig's mother.

- Instead of a specific name designating young animals, the common species name may be used, with or without an additional term referring to newborns. In this case, we have decided to annotate the species name. Since it does not concern information we want to extract, it will not be linked to others entities and thus not included in the resulting database.

We might speculate that also in dog puppies 3'SL aids in orchestration of the gut environment by modulation of specific gut microbiota establishment and mucosal immunity (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, dog is to related to puppies, it is annotated but will not be part of any relations.

Thus, preterm pigs (n = 112) were fed formula without or with HMO supplementation (5–10) g/L … Individual HMO were quantified in colon contents and urine (Rudloff_2019_part-1) (Rudloff et al., 2019)

Preterm pigs indicate that the 112 pigs studied here are newborn and not female. The term pigs are annotated but will not be related to any entities. The colon and urine mentions are clues, indicating that this section of the article speaks about the analysis of the oligosaccharide's effect on newborns.

ix. Names containing species names

In several experiments, authors have used oligosaccharides extracted from animal body parts as standards to determine the pattern of oligosaccharides in another species. In such cases, the origin of the standard is mentioned in the methodology section, specifying the species from which the oligosaccharides were extracted. These species names must be annotated even if they are not linked to other entities.

lacto-Nnovo-pentaose I (LNnP I; Gal(b1–4)GlcNAc(b1– 6)(Gal(b1–3)) Gal(b1 – 4)Glc, N21) and LNnH (Gal(b1 – 4)GlcNAc(b1–6) (Gal(b1 –4)GlcNAc(b1 – 3))Gal(b1 –4)Glc, N24), the latter of which is present in human milk(7), were degraded … by a (b1 – 3/4)-specific galactosidase from bovine testis … The fucosylated structures were identified with the aid of … an a(1 –2/3/4)-specific fucosidase from bovine kidney. (Albrecht_2014_part-2) (Albrecht et al., 2014)

Bovine indicates the species from whom the galactosidase and fucosidase chemicals used to identify oligosaccharides were collected.

x. Qualifier terms

The annotation must not include words that qualify a species without being part of the species name.

In this study, 335 healthy women were recruited in Clifford Hospital, Panyu District, Guangzhou City, Guangdong Province, China, between March 2018 and June 2019. (Liu_2021_part-2) (Liu et al., 2021)

Here, women are qualified as healthy, and this mention is excluded from the annotation.

xi. Human Lewis and secretor type

A link has been established between blood group and milk oligosaccharides in the human species. It has led to the categorization of human individuals into different groups, designated by precise names (Lewis, secretors), based on their genetic pattern. Consequently, several authors use these groups names to refer to human individual. However, we cannot guarantee that these terms will not be used for other mammal species in the future. Because of this ambiguity, we must not annotate these terms.

On postnatal day 180, the total concentration of HMOs in Malawi milk samples from secretors (6.46 +/- 1.74 mg/mL) was higher (P < 0.05) than that in samples from nonsecretors (5.25 +/- 2.55 mg/mL) (Xu_2017_part-1) (Xu et al., 2017)

Secretors and nonsecretors could be considered as indications of the human species but they must not be annotated.

b. Breed

i. Definition

Breed is a subclass of species; thus, it must be related to this entity. The Livestock Breed Ontology organizes breed names by relevant subtrees (https://bioportal.bioontology.org/ontologies/LBO), but only for the following species: buffalo, cattle, chicken, goat, horse, pig and sheep breeds. Each breed annotation belonging to this ontology must be associated with its ID from the livestock breed ontology (ex: the Duroc breed will be associated with the ID “LBO:0000358”). Consequently, each breed name must be annotated, however not all of them will be normalized.

Milk samples from 7 individual dogs were used in the present study: ... One was an Alaskan Husky; Ella (AH). One was a Labrador Retriever (LR). One was a Miniature Schnauzer (SCH). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

In this example, the dog breeds must be annotated even if they are not in the Livestock Breed Ontology.

Porcine milk oligosaccharides (PMOs) were analyzed in six colostrum and two mature milk samples from Dutch Landrace sows. (Difilippo_2016_part-1) (Difilippo et al., 2016)

In contrast, in this sentence, Dutch Landrace belongs to the Livestock Breed Ontology, thus it must be annotated and will be normalized.

ii. Cross breed

When entities follow each other and have independent meanings, we annotate them separately. However, when an entity is defined by the combination of several entities, such as when individuals are born from a crossbreeding, we annotate them altogether. Thus, all mentioned breeds must be annotated as a single entity.

Sows (n = 10) and gilts (n= 7) of a Landrace, Belgian Landrace, Large White and Duroc cross-breed (Sus scrofa) from the Pig Improvement Company (PIC) facility at Grong Grong, N.S.W, Australia were the source for collection of all milk samples. (Wei_2018_part-2) (Wei et al., 2018)

This sentence above clearly states the word cross-breed, thus we must annotate the four breeds mentioned before as one entity.

Three dogs were crosses of Alaskan Husky and German Shorthair Pointer; Annie, Uli and Isis (AH-GP). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The dogs are cross-breed, both breeds are annotated as a single entity.

However sometimes the word cross breed is not mentioned but a symbol in-between the two breeds indicate the species belongs to a cross breed. As with the above example, both breeds must be annotated as one entity.

Saito et al. (1984) described the chemical structure of two neutral disaccharides in colostrum collected 6 h postparturition from Holstein±Friesian cows. (Gopal_2000_part-1) (Gopal & Gill, 2000)

This cow is a crossbreed between Holstein and Friesian, both terms are annotated as one term.

Bovine colostrum from Holstein-Friesian cows and porcine colostrum from Landrace pigs were obtained on-site at Teagasc Food Research Centre, Moorepark (Fermoy, Cork, Ireland) ... while ovine colostrum from Scottish black-faced mountain sheep was kindly donated by Cashel Irish Farmhouse Cheese Makers (Fethard, County Tipperary). (Albrecht_2014_part-2) (Albrecht et al., 2014)

Here, cows are cross-breed while pigs and sheep are from one breed.

c. Geography

i. Definition

This entity indicates where the samples are from. Geographical locations are not always mentioned in the articles; however, they provide relevant information and thus must be annotated when occurring. Location mentions are classified by the geographical database Geonames. Each annotated location must be associated with its ID from Geonames (https://www.geonames.org/).

ii. Precise location

When the laboratory, university, farms, or other structures where samples come from are indicated, they must not be annotated. Those places are too precise and generally not available in Geonames. The information annotated must be at the city, region, state, country or continent levels.

Mature milk from dromedary camels was kindly provided by King Faisal University (Saudi Arabia). (Albrecht_2014_part-2) (Albrecht et al., 2014)

Here, the precise location "King Faisal University" is not annotated, whereas the country name "Saudi Arabia" is.

However, sometimes these precise locations include a geographic name (ex: University of Toulouse). Since all geographic mentions in the text must be annotated, those embedded in precise locations must also be.

Two colostrum samples (7 and 8) were donated by Animal Nutrition Group (Wageningen University) (Table 1) (Difilippo_2016_part-2) (Difilippo et al., 2016)

Wageningen University is located in the city Wageningen in the Netherlands.

Oli6 stage 1 (S1-GIF) and stage 2 goats' milk infant formulas (S2-GIF) and raw goats' milk (pooled milk from a group of ten Saanen goats) were obtained from Nuchev Pty Ltd. (Leong_2019_part-1) (Leong et al., 2019)

When the only information is a precise location, such as the firm above, it must not be annotated and no geographic information will be extracted.

iii. Succession of geographical names

Usually, geographical names following each other designate a place included in the following one. Each name must be annotated as a separate entity.

Bovine colostrum from Holstein-Friesian cows and porcine colostrum from Landrace pigs were obtained on-site at Teagasc Food Research Centre, Moorepark (Fermoy, Cork, Ireland). (Albrecht_2014_part-2) (Albrecht et al., 2014)

Each geographical location, "Fermoy", "Cork" and "Ireland", is annotated as a distinct entity.

It was validated and applied to milk samples from Malawi (88 individuals; 88 samples from postnatal month 6) and the United States (Davis, California; 45 individuals, mean age: 32 y; 103 samples collected on postnatal days 10, 26, 71, or 120, repeated measures included). (Xu_2017_part-1) (Xu et al., 2017)

Likewise, in the sentence below, Davis is a city in California in the United States, all of these geographic entities must be annotated separately.

iv. Abbreviations

Abbreviated geographical names must be annotated.

A decrease (P < 0.05) in HMO concentration was observed during the course of lactation for the US mothers (Xu_2017_part-1) (Xu et al., 2017)

In the following sentence, US is the abbreviation of United States, it must be annotated.

Sows (n = 10) and gilts(n= 7) of a Landrace, Belgian Landrace, Large White and Duroc cross-breed (Sus scrofa) from the Pig Improvement Company (PIC) facility at Grong Grong, N.S.W, Australia were the source for collection of all milk samples. (Wei_2018_part-2) (Wei et al., 2018)

N.S.W stands for New South Wales, a region of Australia

v. Names containing geographical names

Some breed names contain geographical names. These names must not be annotated.

Equine colostrum from Draught foster mares was kindly provided by Coolmore Stud (Fethard, County Tipperary), while ovine colostrum from Scottish black-faced mountain sheep was kindly donated by Cashel Irish Farmhouse Cheese Makers (Fethard, County Tipperary) (Albrecht_2014_part-2) (Albrecht et al., 2014)

"Scottish black-faced" does not mean the samples were from Scotland, it only indicates the breed's name. As a result, it must not be annotated as a geographic entity but as a breed.

vi. Geographic terms used for other purpose

Some geographic locations do not concern the origin of the studied samples. It can be, for instance, the location of the laboratory where the samples were analyzed or the location of the laboratory that has approved the experimentation. This information will be annotated (the algorithm may not be able to exclude those geographical names) but because of the lack of relations with other entities, it will not be included in the database.

All pigs were obtained from the commercial farrowing shed at the Pig Improvement Company facility (Grong Grong, NSW, Australia), as was the source of all milk samples reported in this study. The study protocol was approved by the Animal Care and Ethics Committee of Charles Sturt University, Wagga Wagga, NSW, Australia (13/103). (Wei_2018_part-2) (Wei et al., 2018)

Here the first location is that of the studied individuals and must be annotated. But the second location is from the laboratory that has approved the experimentation. It must be annotated but will not be linked to any entities.

Milk samples were frozen in air tight vials at −20°C or −80°C and subsequently shipped to the Nutrition Laboratory, Smithsonian National Zoological Park, Washington DC, USA. (Taufik_2012_part-2) (Taufik et al., 2012)

In this example, samples have been moved to a laboratory in the USA but they do not come from this place, it is annotated but no relation will be made.

d. Female physiological stage

i. Definition

This entity corresponds to the physiological stage of the studied species, meaning her lactation parity that indicates the number of past gestations. Female physiological stages are organized in the

synonyms list of the FemaleParityThesaurus available at https://doi.org/10.57745/LFXGFO. The terms must be normalized with the first column term of this thesaurus.

ii. Designation of the parity

Several words may be used for the lactation parity. It can be directly stated with words such as primiparous, multiparous, "number"-parity or bred for the "x" time... They are gathered in the FemaleParityThesaurus and must all be annotated.

Milk was collected from second-parity sows (n =3) at farrowing and on days 1, 4, 7, and 24 of lactation. (Tao_2010_part-1) (Tao et al., 2010)

Characterization of porcine milk oligosaccharides over lactation between primiparous and multiparous female pigs (Wei_2018_part-1) (Wei et al., 2018)

Otherwise, the information can be implicit, and included in the species designation. In fact, several names used to designate the species are also indications of the parity of the females. Those words must be annotated as species entities, they are frequently encountered for pigs with "gilts" referring to primiparous females and "sows" to multiparous ones.

Milk samples were collected from gilts and sows on colostrum at day 1, transitional milk at day 3 and mature milk at day 15 of lactation by manual expression during the same time period each morning (0900–1200 h). (Wei_2018_part-2) (Wei et al., 2018)

Sometimes, both female physiological stage and species entities can be found.

Therefore, the structural diversity of PMOs in the milk of female pigs which had been bred at least once (sow) with those which had been bred for the first time (gilt) were assessed in colostrum (day 1), transitional milk (day 3) and mature milk (day 15–21). (Wei_2018_part-2) (Wei et al., 2018)

The pigs (Sus scrofa, Belgian Landrace, Large White, Landrace, and Duroc) including gilts (young female pigs that have not farrowed yet; n = 8) and sows (mature female pigs that have bred at least once; n = 22) were included in this study. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

In this example, authors use the expressions "not farrowed yet" and "bred at least once" to indicate their parity.

iii. Unnecessary information about individuals

Sometimes, authors are giving other information about the individuals, such as age, gestational age, and the number of newborns they had at their pregnancy. This information must not be annotated as illustrated below.

Mid-lactation milk samples (5ml) were collected from a free-living polar bear with one yearling (16 months old) in Svalbard in the Norwegian Arctic. (Urashima_2003a_ part-1) (Urashima et al., 2003a)

The age of the polar bear is not a relevant information.

The inclusion criteria were that the pregnant women had lived in the area for more than two years, were aged 20–35 years, planned to breastfeed for more than 3 months, had singleton pregnancies, and had a gestational age of 37–42 weeks. (Liu_2021_part-2) (Liu et al., 2021)

None of this information concerns the lactation parity, we must not annotate it.

e. Individual number

i. Definition

This entity gives information on the number of individuals that have been studied for the experiment. No normalization is needed.

ii. Number of samples

The number of individuals must not be confused with the number of samples. The number of samples must not be annotated. In fact, sometimes several samples have been collected from one individual.

A total of 488 breast milk samples were collected from 335 healthy mothers at five different time points (Liu_2021_part-1) (Liu et al., 2021)

In this sentence, we only annotate the number of individuals.

Eight milk samples from DutchLandrace sows were collected ... Two colostrum samples (7 and 8) were donated by Animal Nutrition Group (Wageningen University) (Table 1), while six samples, of which 4 were colostrum and 2 were mature milk samples, matching the corresponding colostrum samples, were obtained from Proefaccommodatie de Tolakker (Utrecht University, Utrecht, The Netherlands). (Difilippo_2016_part-2) (Difilippo et al., 2016)

Here, we must not annotate “eight” since it is the number of milk samples, not the number of sows. Later in the sentence, this number is divided between the lactation stage. We have decided not to annotate this information. First of all, because it designates the sample, and secondly because we only annotate the global number of individuals.

iii. Numbers expressed in letters

The number of studied individuals can be written in letters as well.

Urashima et al. [38–40] used samples from two polar bears, four Japanese black bears and only one from an Ezo brown bear. (Urashima_2001_part-1) (Urashima et al., 2001)

iv. The global number of individuals

When the number of studied individuals is subdivided between different groups of individuals, we must extract all those numbers and not only the global number of individuals. Later on, the relations will only be made with the numbers for which information is available.

Greater Galago (Galagidae: Otolemur crassicaudatus): 17.3 mL milk was obtained by pooling milk from four females. Three females were milked ... at 19, 19 and 96 days postpartum, and the fourth female was milked ... at 44, 59 and 75 days postpartum. (Taufik_2012_part-2) (Taufik et al., 2012)

In this article, the authors have divided females based on the postpartum age. Consequently, they give the precise number of females studied in each group. We must annotate all the numbers mentioned. Likewise, the indications about the females (such as postpartum age) must all be annotated to indicate the different types of samples the study gives information about. However, fourth is a rank and not a number, thus it must not be annotated.

A total of 488 milk samples were collected from 335 healthy lactating mothers: 96 at 0–5 days, 96 at 10–15 days, 104 at 40–45 days, 100 at 200–240 days, and 92 at 300–400 days postpartum. (Liu_2021_part-2) (Liu et al., 2021)

Likewise, in this example we must annotate all the numbers of women that participated in the study. Likewise, all postpartum age indications must be annotated.

Mongoose Lemur (Lemuridae: Eulemur mongoz): 9.4 g milk was collected from one female ... at 25, 36 and 62 days postpartum, from a second female... at 30, 41, 58 and 81 days postpartum, and from a third female ... at 56 days postpartum. (Taufik_2012_part-2) (Taufik et al., 2012)

In the sentence above, each female is cited individually. Yet, we are only annotating numbers, and not ranks. Thus, we must annotate “one” but not the mentions “second” and “third”.

Milk samples from 7 individual dogs were used in the present study: Three dogs were crosses of Alaskan Husky and German Shorthair Pointer ... One was an Alaskan Husky – English pointer cross ... One was an Alaskan Husky ... One was a Labrador Retriever (LR). One was a Miniature Schnauzer (SCH). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, each breed of dogs has a specific number of individuals, which are all mentioned. All of these indications must be annotated.

The OS 3’SL, 6’SL and 2’FL were quantified in milk samples of 3 Alaskan huskies (AH, AH-GP, AH-EP), 1 Labrador retriever (LR) and 1 Schnauzer (SCH) using external standard curves of authentic standard compounds. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Likewise, in the following sentence, we must annotate the number of each dog breed where the oligosaccharides have been found.

Complete data were available for all 3 time points in 25 pigs, including 18 sows and 7 gilts: mature milk samples were not collected from 4 sows and 1 gilt because they were culled before 15 d of lactation due to poor milk production. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

Here, gilts and sows are species names (more detailed names than pigs). Thus, they designate the studied individuals. Therefore, the number of individuals from each of these species needs to be annotated.

v. Number of species included in a larger group

Sometimes, the numbers of species from an order, a family or another type of classification group are indicated. They must not be annotated, because we do not know how many individuals from these specific species (belonging to the order, family, or suborder...) have been studied.

The four strepsirrhine species varied considerably in the numbers of individual milk oligosaccharides. (Taufik_2012_part-2) (Taufik et al., 2012)

This sentence only states that among the species belonging to the strepsirrhine suborder, four have been studied. It must not be confused with the number of individuals.

vi. N for the number

Sometimes the number is preceded by the indication "n =", we must annotate the number only.

The pigs (Sus scrofa, Belgian Landrace, Large White, Landrace, and Duroc) including gilts (young female pigs that have not farrowed yet; n = 8) and sows (mature female pigs that have bred at least once; n = 22) were included in this study. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

3. Entities referring to samples

a. Sample type

i. Definition

This entity refers to the samples used during the experimentation. Most of the time these samples are mother's milk. However, some documents contain particular designations for these samples. To overcome this issue, sample type names are organized in the synonyms list of the SampleThesaurus at https://doi.org/10.57745/LFXGFO. The terms must be normalized with the first column term of this thesaurus.

ii. Pooled samples

When there is an indication that samples have been pooled (such as the terms "pool" or "mix"), this indication must be annotated within the sample type entity. Usually, the terms "pooled" or "mixed" are linked to the sample type as adjectives and thus both words (milk and indication of pooling) must be annotated as one.

Raw goats' milk (pooled milk from a group of ten Saanen goats) were obtained from Nuchev Pty Ltd. (Leong_2019_part-1) (Leong et al., 2019)

But sometimes, they are further away in the text, and both words must be annotated separately.

7.0 mL milk was obtained by pooling samples from 4 females (Taufik_2012_part-2) (Taufik et al., 2012)

Generally, indications of pooling are close to the word "samples" or "milk".

The samples were pooled and frozen at −80 °C. (Lee_2016_part-1) (Lee et al., 2016)

Given a weaning age of 4.4 months (Table 1), this pooled sample represents early to mid-lactation. (Taufik_2012_part-2) (Taufik et al., 2012)

Five samples from the same breed were mixed together to form a sample to analyse. (Cheng_2016_part-2) (Cheng et al., 2016)

iii. Young individual samples

Several articles study the impact of the oligosaccharides found in female milk on the newborn. Consequently, they analyze samples from newborns, such as colon content, feces, and urine. These samples should not be annotated since they do not reflect the exact composition of oligosaccharides found in the mother's milk.

Individual fecal samples were collected from three piglets, 1–2 d and 1 w old (Proefaccommodatie de Tolakker), within 24 h after porcine milk ingestion. (Difilippo_2016_part-2) (Difilippo et al., 2016)

In this article, the authors analyze mothers' milk composition to determine the oligosaccharide pattern. Then, they analyze the oligosaccharide content in fecal samples from piglets to understand how they are digested. Consequently, even though these fecal samples are giving an oligosaccharide pattern, they must not be annotated since they have undergone modifications in comparison to the one from the mother's milk.

iv. Industrial samples

Industrial food oligosaccharide composition must not be annotated. It can be yogurt, cheese, or dietary supplement... Sometimes, one article can analyze both milk samples from individuals and industrial samples. Thus, it is necessary to only annotate information regarding milk.

3'SL was the most abundant oligosaccharide in bovine whey permeate, followed by 3'sialyl-galactosyl-lactose, 6'SL(Neu5Ac), 6'SLN(Neu5Ac), triose B, N-acetylglucosaminyl-lactose, and sialyllacto-N-tetraose (LST) in order of descending abundance (Table 3b). (Lee_2016_part-1) (Lee et al., 2016)

Whey permeate is a by-product of cheese, thus even if it comes from bovine milk, it has undergone several transformations and does not reflect the exact composition of bovine milk. Bovine and oligosaccharides found in this industrial product will be annotated, but not linked to a sample entity.

Infant formula also belongs to industrial samples and must not be annotated. The algorithm must be trained not to annotate the word milk when it is surrounded by the mention of "infant formulas".

Oli6 stage 1 (S1-GIF) and stage 2 goats' milk infant formulas (S2-GIF) and raw goats' milk (Leong_2019_part-1) (Leong et al., 2019)

In this article, authors have studied raw milk, directly extracted from goats as well as goats' milk infant formulas, that are industrially processed. The only sample we must annotate is the raw milk.

v. Detailed milk extraction method

Frequently, authors give details about the method they have used to extract milk. This information is excluded.

As mouse milk was difficult to collect directly from lactating mouse by squeesing, mammary tissue was used for extraction of milk oligosaccharides. After sacrifice, the entire mammary gland of maternal mouse was gently peeled off with a scalpel, and then immersed in phosphate-buffered saline (PBS) until the white milk was extracted completely. (Li_2021_part-1) (Li et al., 2021)

In this article, mouse's milk has not been extracted as in the other animals. Therefore, the authors explain the method, which include mammary tissue. It must not be annotated as sample, only the milk that has been obtained from this tissue must be annotated.

Unlike dairy cows, the porcine mammary gland does not have cisternae in which to store milk. Milk collection can only be carried out after inducing the milk ejection reflex through release of oxytocin induced by the suckling stimulus of piglets for at least 1 min. Sows nurse piglets approximately 20 times per day, and the milk flow lasts for about 10 to 20 s. In this study, 2 to 5 mL of milk was collected from each gilt and sow after the milk letdown was induced by the suckling stimulus from piglets during the 10- to 20-s window of opportunity. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

Many details are given about the milk test, they must not be annotated.

b. Lactation stage

i. Definition

This entity gives information on the mothers' milk. In fact, during lactation, milk characteristics evolve. Thus, we generally distinguish 3 to 4 types of milk based on the period of lactation, usually called colostrum, transition milk, mature milk and weaning. The list of the commonly used term is available in the LactationStage-concepts file at https://doi.org/10.57745/LFXGFO. Depending on the sentence, terms can occur as such or can be followed by "milk", which must be annotated with the terms (Warning: sometimes "milk" is not directly adjacent to the word, and the annotation will be discontinuous). These terms are typical of the species studied (species do not have the same duration of lactation, consequently general terms such as early lactation may hold different meanings across species), that is why no normalization is made. The database will give raw terms, the user needs to know the species characteristics to link these terms to a precise period.

Milk from 8 gilts and 22 sows was collected at 3 stages of lactation (colostrum, transition, and mature milk). (Jahan_2016_part-1) (Jahan, Wynn & Wang, 2016)

In this example, 3 terms are used: colostrum, transition and mature milk. Mature is the only term adjacent to "milk", consequently milk is annotated with it (and as a sample type entity). Transition is in a discontinuous annotation with milk.

Up to 250 mL of milk were collected from seven sows at each of three time points: day 0 (colostrum), days 7–9 (mature), and days 17–19 (weaning). Colostrum was collected within 6 h of farrowing and 3-

day intervals were used for mature and weaning milk to ensure representative sampling. (Mudd_2016_part-1) (Mudd et al., 2016)

In this example, we can see the variety of terms used to describe the milk. Both examples are available in this sentence: terms as such or followed by milk (discontinuous annotation with mature milk).
As we will see in the next paragraph, farrowing does not describe milk but a period.

ii. Unusual words

Different terms can be used to characterize the type of milk considered. Sometimes, the terms used are not directly linked to the milk but indicates the period of lactation.

The figure also shows that all major peaks decrease significantly from farrowing to midlactation and some increase at late lactation. (Tao_2010_part-2) (Tao et al., 2010)

Indeed, fucosylated PMO increased during lactation, mirroring a similar trend observed for neutral and type I OS content during early lactation. (Salcedo_2016_part-1) (Salcedo et al., 2016)

In the present study we detected the trisaccharide 2'-FL in early lactation milk from an aye-aye with a sick infant that subsequently died; it was not found in mid-lactation milk of this species. (Taufik_2012_part-2) (Taufik et al., 2012)

Those words describing a period can be followed by "milk" which must also be annotated (as well as a sample type entity).

Furthermore, HMO profiles changed in later lactational stages. With a prolonged sampling time, we found that 3-FL levels exceeded 2'-FL levels in the late lactation period and became the predominant HMO compared with other HMOs. [...] 6'-SL was high in breast milk at the beginning of lactation, and its level at 10–15 days postpartum was 26 times higher than that at 300–400 days postpartum. (Liu_2021_part-2) (Liu et al., 2021)

Here are some other terms encountered in articles.

Milk was collected at eight different time points throughout the lactation period from day 0 (birth) to day 40 (on days 0, 1, 3, 5, 10, 20, 30, 40). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Farrowing refers to the birth period, it must be annotated as a lactation stage. It indicates that the author is talking about the first milk.

iii. Synonyms

The words can take several forms. In the following sentence, colostrum is used as an adjective (colostral), it must be annotated.

3'-SL is also the most abundant oligosaccharide in cow colostrum (about 49% of the total colostral BMOs), while it is a minor component in human colostrum. (Difilippo_2016_part-2) (Difilippo et al., 2016)

iv. Used as sample type

Frequently, authors refer to lactation period in the same way as sample type (milk, pooled milk...), especially for colostrum which is the name of the first milk produced by the mother. It must, however, still be annotated as a lactation period entity.

HPLC chromatograms of PMO from colostrum and milk collected at farrowing and days 1, 4, 7, and 24 of lactation from the same sow are shown in Figure 3. (Tao_2010_part-2) (Tao et al., 2010)

In this sentence, colostrum and milk are clearly distinguished as two types of samples. Still, they must be annotated as two different types of entity.

c. Post-partum age

i. Definition

This entity indicates the duration since the mother has given birth. It is strongly related to the Lactation stage since they both give information on the samples. Consequently, we can find them both in the same sentence. No thesaurus will be used, but these entities have to be standardized using days as unit of time.

ii. Formulation

Post-partum age can be expressed in several ways. Generally, the number of days after the mother has given birth is used with patterns such as: “x days postpartum”, “x days after delivery”, “x days after farrowing”. However, it can also be formulated with reference to milk secretion, such as “day x of lactation” and “x days of lactation”. Globally, all these expressions are used to designate birth as the starting date.

Milk samples were collected from gilts and sows on colostrum at day 1, transitional milk at day 3 and mature milk at day 15 of lactation. (Wei_2018_part-2) (Wei et al., 2018)

Milk samples were collected from colostrum, days 1, 5, 10, and 21 after farrowing. (Cheng_2016_part-2) (Cheng et al., 2016)

This example shows discontinuous annotations: days with 5, 10 and 21 as well as after farrowing with days 1, 5 and 10.
Farrowing belongs to two entities here: lactation stage but also post-partum age.

The evolution of PMO was investigated in the milk from 3 healthy sows at prefarrowing, farrowing, and d 7 and 14 postpartum by Nano-LC Chip Quadrupole-Time-of-Flight mass spectrometer. (Salcedo_2016_part-1) (Salcedo et al., 2016)

Sometimes, “days” is abbreviated by “d”, it must be annotated. Furthermore, in this example “d7” is in a discontinuous annotation with “postpartum” and “14 postpartum” is in a discontinuous annotation with “d”.

Milk was collected at eight different time points throughout the lactation period from day 0 (birth) to day 40 (on days 0, 1, 3, 5, 10, 20, 30, 40). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, there are no indication of a starting point. Consequently "day" must be interpreted as the age of the individual and not as the days since a starting date (postpartum, farrow, beginning of lactation...)

iii. Ambiguous period

When the days are not stated precisely (end of lactation, first days...), the mention must not be annotated as post-partum age but as lactation stage. In fact, among species, those hazy terms may not signify the same period (first days of lactation last longer for species whose lactation is quite long).

In the first days of lactation it was present at levels around 7.5 g/L (12 mM), which dropped rapidly to approximately 1.5 g/L at 10 days of lactation. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here first day of lactation is too confusing, whereas 10 days of lactation gives the precise period.

Especially during the first week postpartum the high 3'SL concentration and 3'SL to lactose ratio might indicate a specific physiologic need of the newborn puppies. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The other dog breeds had lower levels, yet they also had a peak 2FL level (0.3–0.5 g/L) during the first 5 lactation days followed by a drop within the first 10 days to around 0.1 g/L. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, authors give precise days, thus we must annotate them.

iv. Exact date

When the exact dates of sampling are indicated, they must not be annotated, as this would mean calculating the number of days since farrowing.

Greater Galago (Galagidae: Otolemur crassicaudatus): Three females were milked on 21 May 1990, 24 May 1990, and 9 February 1990 at 19, 19 and 96 days postpartum, and the fourth female was milked on 18 December 1989, 3 January 1990, and 19 January 1990, at 44, 59 and 75 days postpartum. (Taufik_2010_part-2) (Taufik et al., 2012)

Both milking dates and postpartum age are given in this sentence. We must annotate only the postpartum age.

v. Time interval

When post-partum age is divided into intervals, each one must be annotated.

In this study, 335 healthy women were recruited in Clifford Hospital, Panyu District, Guangzhou City, Guangdong Province, China, between March 2018 and June 2019. Breast milk samples were collected

at five different time points: 0–5 days, 10–15 days, 40–45 days, 200–240 days, and 300–400 days post-partum. (Liu_2021_part-2) (Liu et al., 2021)

vi. Confusion of postpartum age with lactation stage

Confusion must not be made with the entity Lactation stage.

To illustrate the relative abundances of the major OS species in porcine milk, three milking points were chosen to represent early (farrow), mid (day 4), and late lactation (day 24) (Tao_2010_part-2) (Tao et al., 2010)

Here we can see the narrow boundary between lactation stage and postpartum age. They complement each other, postpartum age gives more detailed information about the lactation stage (those stages differ between species; thus, the days are given to reduce the ambiguity). However, farrow must be annotated as a lactation stage and not as a postpartum age entity since it does not refer to a period or duration after the birth.

For oligosaccharide identification and characterization, samples from one dog (Annie) were selected from the initial, middle and end stage of lactation (day 1, day 10 and day 40). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Likewise, here postpartum age gives detailed information about lactation stage.

HPLC chromatograms of PMO from colostrum and milk collected at farrowing and days 1, 4, 7, and 24 of lactation from the same sow are shown in Figure 3. (Tao_2010_part-2) (Tao et al., 2010)

This sentence gathers the two entities. Firstly, colostrum and milk are stated as two different milk samples, afterwards terms indicate the precise period when these samples were collected. Nonetheless, they must not all be annotated as postpartum age.

Results from this study show that lactating women continue to provide their offspring with a high level of 2'-FL one year after delivery, suggesting that 2'-FL may play an important role for infants in early life. (Liu_2021_part-1) (Liu et al., 2021)

One year after delivery is a long-term entity, but it must be annotated and transformed into days. It indicates that an oligosaccharide is still present in the human milk one year after the first milk.

Saito et al. (1984) described the chemical structure of two neutral disaccharides in colostrum collected 6 h postparturition from Holstein±Friesian cows. (Gopal_2000_part-1) (Gopal & Gill, 2000)

Here, the distinction between postpartum age and lactation stage is clear. Lactation stage indicates the type of milk collected, while postpartum age describes the period to which this milk refers.
The entity still has to be standardized using the day as unit of time instead of the hour.

d. Methodology of analysis

i. Definition

The study methodology refers to the method used to measure oligosaccharide quantity and quality. The study methodology is normalized using the MOMethodThesaurus available at https://doi.org/10.57745/LFXGFO. The attribute ID must be associated with the corresponding methodology entity.

ii. Abbreviations

Frequently, authors use abbreviations of the methodology name. This abbreviation must be annotated and when both the entire name and the abbreviation follow each other, they must be annotated separately.

A high-performance anion-exchange chromatography-pulsed amperometric detector (HPAEC-PAD) was used to quantify 2'-fucosyllactose (2'-FL) ... (Liu_2021_part-1) (Liu et al., 2021)

iii. Detailed methodology information

Information explaining how the laboratory experimentation have been conducted must not be annotated, we only want the methodology name.

A high-performance anion-exchange chromatography with pulsed amperometric detector (HPAEC-PAD) was used to measure HMOs in breast milk ... Briefly, a 200 µL homogenized milk sample was diluted 10 times with Milli-Q water ... and then vortexed for 1 min to mix thoroughly. The mixture was filtered with a 0.22 µm nylon filter to remove proteins and lipids ... The identification of HMOs was performed on a Thermo Fisher HPAEC ICS 5000 series (Thermo, Waltham, MA, USA) system equipped with a separation column (CarboPacTM PA1, 4 × 150 mm, Thermo) connected with a guard column (CarboPacTM PA1, 4 × 50 mm, Thermo). (Liu_2021_part-2) (Liu et al., 2021)

4. Entities referring to oligosaccharides

Oligosaccharides are a succession of several sugar molecules called monosaccharides linked to each other thanks to glycosidic bonds. Monosaccharides take a cyclic form by condensation between aldehyde function (-CHO) and alcohol function (-OH) (Figure 1). This phenomenon is called intramolecular hemiacetalisation. Aldehyde function located at carbon 1, called anomeric carbon, is no longer free, we obtain an -OH hemiacetal with a reductive function. Based on anomeric carbon configuration, we distinguish 2 positions (α or β) for -OH hemiacetal (Varki et al., 2022).

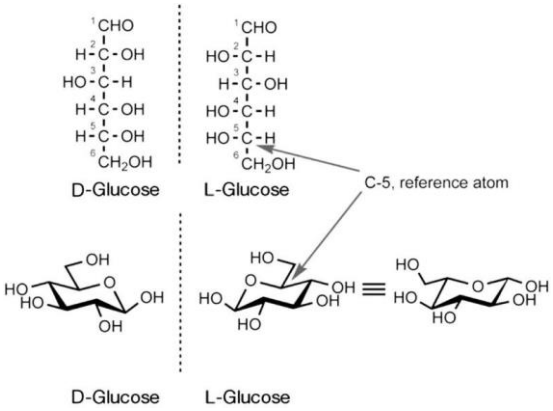

Figure 1: Monosaccharide intramolecular hemiacetalisation leading to α or β configuration for -OH hemiacetal of the anomeric carbon (Varki et al. 2022)

During oligosaccharides synthesis, bonds always occur between -OH hemiacetal of one monosaccharide with

- -OH from alcohol function of the other monosaccharide (bond C1 with C2, C3, C4 or C6), whose hemiacetal function remains free
- -OH hemiacetal (bond C1 à C1)

Nomenclature always goes from left to right, monosaccharide residue engages his anomeric carbon in the glycosidic bond with one carbon from the monosaccharides located at his right. Thus, based on this last position, nomenclature will be different: "α or β – monosaccharide residue (1-> n (number of the C other than anomeric)) - α or β - monosaccharide residue" (Varki et al., 2022).

More specifically, milk oligosaccharides are only composed of 5 monosaccharide types outlined thereafter according to the Symbol Nomenclature for Glycans (SNFG) (Varki et al., 2015) (Figure 2):

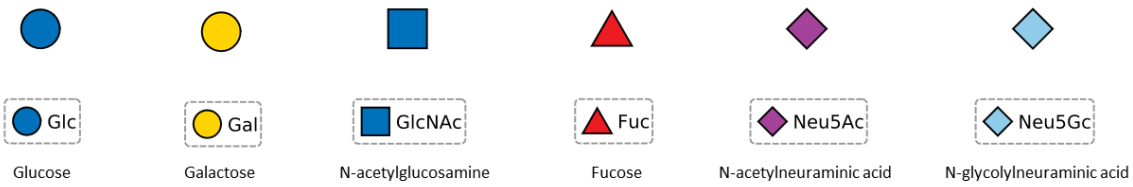

Figure 2: Monosaccharides included in the composition of milk oligosaccharides (Varki et al., 2022)

N-glycolylneuraminic (NeuGc) is a sialic acid, such as N-acetylneuraminic (NeuAc), however humans are not able to synthesise NeuGc. On the contrary, non-primate mammals synthesise both NeuGc and NeuAc, thus several species have more oligosaccharides containing NeuGc than NeuAc (Sprenger et al., 2022).

Oligosaccharides core is always composed as followed: lactose (Gal(β1-4)Glc) (Figure 3)

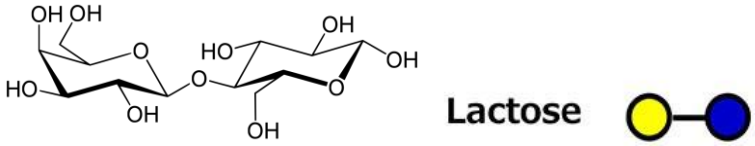

Figure 3: Structure of the disaccharide lactose: Gal(β1-4)Glc (Ujihara & Kentaro, 2022)

or N-acetyllactosamine (Gal(β1-4)GlcNAc) in the reducing end, which is the right part of the molecule when represented on a plan (Figure 4) (Balogh, Jankovics & Béni, 2015).

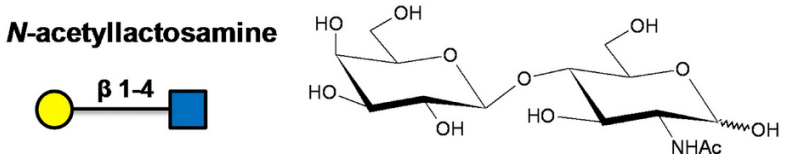

Figure 4: Structure of the disaccharide N-acetyllactosamine : Gal(β1-4)GlcNAc (Balogh, Jankovics & Béni, 2015)

Considering N-acetyllactosamine structure, 2 structure types can be added (Figure 5):

- Galactose β linked by his anomeric carbon to carbon 3 of N-acetylglucosamine with Lacto-N-biose (Gal(β1-3)GlcNAc): **type I oligosaccharides**
- Galactose β linked by his anomeric carbon to carbon 4 of N-acetylglucosamine with N-acetyllactosamine (Gal(β1-4)GlcNAc): **type II oligosaccharides** (Bode, 2012)

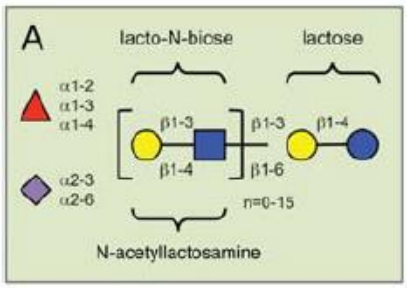

Figure 5: Representation of the two structures types encountered in milk oligosaccharide structures (Bode, 2012)

Based on linkage type between lactose and lacto-N-biose or N-acetyllactosamine, we distinguish:

- Branched structures in *iso*-oligosaccharides when this is a β1-6 bond between the disaccharides
- Branched structures in *para*-oligosaccharides when this is a β1-3 bond between the disaccharides

Other monosaccharide types can be connected to the core molecule (Figure 6):

- Fucose in α1-(2, 3 or 4)
- Sialic acid (NeuGc or NeuAc) in α2-(3 or 6) (Sprenger et al., 2022)

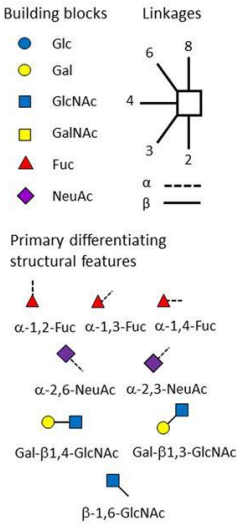

Figure 6: Representation of the linkages between monosaccharides that can occur in the milk oligosaccharide's structure (Sprenger et al., 2022)

According to the chemical composition of oligosaccharides, we classify them into distinct categories (Figure 7):

- **Acidic milk oligosaccharides:** contain one or more sialic acid in their structure
  - **Fucosylated acidic milk oligosaccharides**: contain one or more fucose
  - **Non fucosylated acidic milk oligosaccharides**: do not contain any fucose
- **Neutral milk oligosaccharides**: do not contain any sialic acid in their structure
  - **Fucosylated neutral milk oligosaccharides**: contain one or more fucose
  - **Non fucosylated neutral milk oligosaccharides:** do not contain any fucose (Thurl et al., 2017)

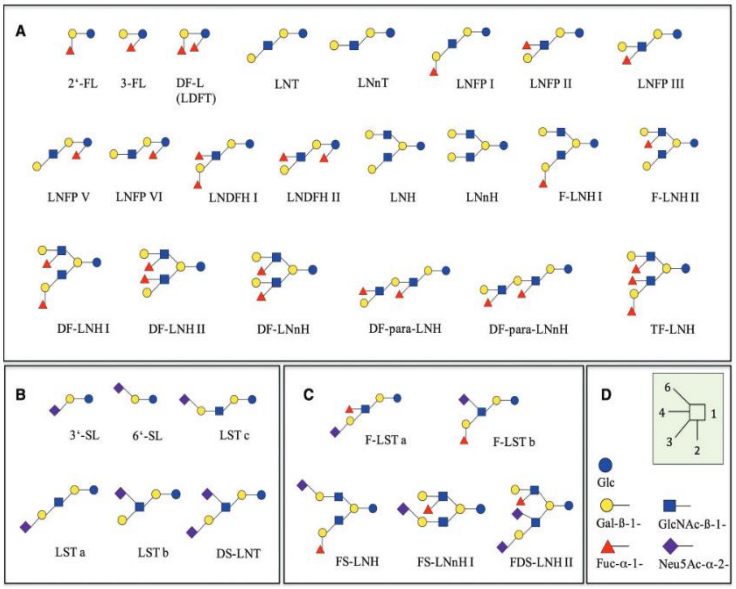

Figure 7: Classification of several oligosaccharides. Those in the A classification are neutral fucosylated, those in the B group are acidic oligosaccharides, those in the C group are acidic fucosylated (Thurl et al., 2017)

Based on this classification, we have decided to annotated both oligosaccharide names and several groups including multiple molecules (in bold letters). These two properties are contained in distinct entities.

a. Oligosaccharide names

i. Definition

This entity gathers the molecule names. Depending on the authors, these molecules can be expressed in very different ways. Thus, oligosaccharide names frequently found in articles are listed in the MilkOligoThesaurus at https://doi.org/10.57745/RA5DAC (Rumeau et al., 2024). The first column of this thesaurus must be used to normalize the molecule name, when this last is empty (several molecules have not been named yet), the second column corresponding to the abbreviation names must be used instead.
Additionally, the different specificities frequently encountered are listed below.

ii. Term oligosaccharides (OS)

The term “oligosaccharides” is not a molecule name, it gathers all the oligosaccharides. However, it can be included in different situations: when not surrounded by any other terms designating

oligosaccharides or when surrounded by oligosaccharides (this part will be explained in the entity Oligosaccharide type).

When it is not surrounded by any other terms designating oligosaccharides, we distinguish two cases:

- If it is surrounded by a number, it must be annotated, since it indicates the total number of oligosaccharides found in the species studied.

Brown capuchin colostrum contains at least six oligosaccharides (Table 1), five of which are also found in human milk. (Urashima_2001_part-1) (Urashima et al., 2001) No molecule names are mentioned here, however, the number six gives us information on the total number of oligosaccharides found in the species.

- Otherwise, when no number is mentioned, it must not be annotated since it does not give any information.

Ovine, caprine and equine colostrum oligosaccharides are also shown in Table 1. (Urashima_2001_part-1) (Urashima et al., 2001)

Sometimes, the term oligosaccharide is abbreviated by "OS" or even "MO" for milk oligosaccharides, it must also be annotated.

Approximately 29 different OS were identified in porcine milk, compared to 40 in bovine milk (9) and >130 in human milk (3). (Tao_2010_part-2) (Tao et al., 2010)

iii. Abbreviations

The abbreviation "species MO" must be annotated. It is frequently used for Human, Bovine, and Porcine species, as HMO, BMO, and PMO respectively. When the long form of an abbreviation is used, it is annotated as an oligosaccharide name and smaller entities embedded inside it must also be annotated with their respective entity types (e.g., species and sample type entities).

The composition of porcine milk oligosaccharides (PMO) was analyzed during early lactation and their relation to piglet gut microbiome was investigated. (Salcedo_2016_part-1) (Salcedo et al., 2016) Porcine and milk are annotated with their respective entity types. The abbreviation PMO is only annotated as an Oligosaccharide name.

iv. Lactose

Lactose is a disaccharide and an important component of the milk of several mammals. However, it does not belong to milk oligosaccharides, it is part of the oligosaccharide structures that, as we have stated before, are completed by at least one monosaccharide.

The polar bear colostrum contained 3'-N-acetylneuraminyllactose, A-tetrasaccharide, 2'-fucosyllactose and lactose (Urashima_2003a_part-2) (Urashima et al., 2003a)

v. General terms

Other general terms must not be taken into account. They are used to characterize sugars more generally.

Comparing this finding with literature reporting on trisaccharides in cow, goat, sheep, and horse milk, the trisaccharide in porcine milk was assigned putatively to Gal(α1-3)Gal(β1-4)Glc. (Difilippo_2016_part-2) (Difilippo et al., 2016)

Trisaccharide does not represent a type of oligosaccharides. It includes all sugars containing 3 monosaccharides; thus, it is not annotated. Conversely, Gal(α1-3)Gal(β1-4)Glc designates a specific oligosaccharides which must be included.

Bovine milk normally contains 1–2g.L of free saccharides other than lactose but larger amounts occur in colostrum; the content for milk is significantly lower than that for human milk. (Urashima_2001_part-1) (Urashima et al., 2001)

Likewise, saccharides are a larger group than oligosaccharides and must not be annotated.

Based on phenol-sulfuric acid analysis of whole milk, the total saccharide content of mid-lactation milk are: greater galago 6.4%, aye-aye 5.8%, Coquerel's sifaka 4.1% and mongoose lemur 7.9% ... Thus strepsirrhine milks are tentatively estimated to contain about 2.5%, 1.5%, 1.0% and 2.0% oligosaccharide in greater galago, aye-aye, sifaka and mongoose lemur, respectively. (Taufik_2012_part-2) (Taufik et al., 2012)

This sentence illustrates the aforementioned specificities. Percentages of total saccharide are bigger than those of oligosaccharide, meaning total saccharide is a larger group and must not be annotated.

vi. Isomers

Isomers are oligosaccharides with the same chemical formula but different arrangements of atoms. The different isomeric forms have similar monosaccharide compositions but differ in a linkage or in the position of a monosaccharide (Soult 2022). General terms are used to gather those isomeric forms with the same formula: fucosyllactose includes 2 molecules namely 2'-Fucosyllactose and 3'-Fucosyllactose, they are both composed by the same monosaccharides but their location differs. Frequently, the technology used to analyze oligosaccharides is not able to distinguish between isomers, thus authors indicate the presence of the isomers groups without specifying the oligosaccharide. Terms that refer to isomers must be annotated. Those terms are available in the MilkOligoThesaurus at https://doi.org/10.57745/RA5DAC (Rumeau et al., 2024).

In milk of a mongrel dog the formation of the OS fucosyllactose, 3'sialyllactose and 6'sialyllactose was reported and in milk of a beagle dog 2'fucosyllactose and sialyllactose were identified together with sulphated lactose. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The example illustrates this phenomenon with fucosyllactose (with the isomer 2'-Fucosyllactose cited later in the sentence) and sialyllactose (with 2 of the 3 isomers stated above).

Acidic oligosaccharides, including sialyllactose (SL) and sialyllactosamine (SLN), are predominant, and also a large portion of neutral tri- and tetra-saccharides is observed. (Lee_2016_part-1) (Lee et al., 2016)

Likewise, here SL can be 3'SL or 6'SL, which is also the case for SLN. However, the general terms tri and tetra-saccharides must not be annotated (see the general terms paragraph above).

Nine neutral oligosaccharides have been described in bovine colostrum (Table 1) [7,11,12], including four galactosyllactoses which are found in the mature milk as well [13]. (Urashima_2001_part-1) (Urashima et al., 2001)

The isomer galactosyllactose is included. This example as well as the previous one illustrates the confusion that must not be made between isomers and the oligosaccharide types. The latter are detailed in a subsequent section.

vii. Doubt about the exact molecule

Methods to identify milk oligosaccharides may not always allow the experimenter to determine the exact structure of the molecule. When it is the case, the article can give isomer groups as seen before, or it gives several molecules that could correspond to the information obtained from the experimentation. All the candidates must be annotated.

We considered two possible structures, Gal(β1-3)GlcNAc(β1-3) Gal(β1-4)[Fuc(α1-3)]Glc (lacto-N-fucopentaose V; LNFP V, type I) and Gal(β1-4)GlcNAc(β1-3)Gal(β1-4) [Fuc(α1-3)]Glc (type II).

Due to the proximity of elution times and the lack of MS-grade standards, it was not possible to distinguish among LNH/LNnH and LNT/LNnT isomers by Nano-LC Chip QTOF MS. (Mudd_2016_part-2) (Mudd et al., 2016)

In this article, the authors clearly indicate their inability to distinguish between two forms. The two must be annotated.

viii. Composition

As mentioned above, isomers are similar molecules (same monosaccharides composition) that differ with their linkages. Thus, composition, which indicates the monosaccharides composing the molecules and their number, are the same among isomers and must be annotated. This composition is indicated in the MilkOligoThesaurus in the form of Hex_HexNAc_Fuc_Neu5Ac_Neu5Gc (Hex: Hexose (Glucose or, Galactose); HexNac: N-acetylgalactosamine; Fuc: Fucose; Neu5Ac: N-acetylneuraminic acid; Neu5Gc: N-glycolylneuraminic acid) with the words replaced by numbers such as 2_1_0_0_0. The other composition form listed in the MilkOligoThesaurus is Hex(n)HexNAc(n)dHex(n)NeuAc(n)NeuGc(n), with n being the number of monosaccharides such as Hex(2)HexNAc(1) (here you can see that whenever a monosaccharide is absent its name is not mentioned, unlike the previous composition form where it is associated with a 0) (Rumeau et al., 2024).

In particular, five OS compositions, including isomers of the bifidogenic Sialyllactose (composition 2_0_0_1_0; SL), the Lacto-N-Tetraose series (composition 3_1_0_0_0: lacto-N-tetraose (LNT) and

lacto-Nneotetraose (LNnT)) as well as the Lacto-N-Hexaose series (composition 4_2_0_0_0; lacto-N-hexaose (LNH) and lacto-N-neohexaose (LNnH)) were detected in all the samples alongside a trisaccharide (3_0_0_0_0) and a pentasaccharide (4_1_0_0_0, likely lacto-N-pentaose (LNP)). (Trevisi_2021_part-1) (Trevisi et al., 2020)

The authors clearly indicate that they have detected isomer forms belonging to a specific group of oligosaccharides and give the name, the abbreviations as well as the composition of the generic OS group. In this study, authors were not able to separate the different isomeric forms of the OS. As already mentioned, trisaccharide and pentasaccharide are too general and are not annotated.

Nonetheless, there are no consensus among authors, and this composition can be written differently. Below are the most frequently encountered forms:

For example, with BMO, peak 1 was determined to be lacto-N-neotetraose (LNnT) with m/z 710.275 [3Hex + 1HexNAc + H]+ . (Tao_2010_part-2) (Tao et al., 2010)

This form is similar to Hex(n)HexNAc(n)dHex(n)NeuAc(n)NeuGc(n), but the numbers precede the monosaccharide abbreviations, which are separated by a plus sign.

In canine milk we identified a somewhat exotic milk oligosaccharide with an m/z ratio corresponding to the tetrasaccharide Hex2-HexNAc-dHex. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

Here, the structure is similar to Hex(n)HexNAc(n)dHex(n)NeuAc(n)NeuGc(n), but with no brackets and when there is only one monosaccharide present no number is mentioned (instead of 1).

The [M H] of peaks #3 and #5 are identical at m/z 632.2, with the composition of Hex2Neu5Ac1 (H2S1), and these two can be tentatively assigned as the two isomeric sialylated lactose (SL) widely found in mammalian milk as the main components. (Li_2021_part-1) (Li et al., 2021)

This extract gathers several oligosaccharides name annotations. The Hex2Neu5Ac1 is the composition name such as H2S1 with H, Hex; N, HexNAc; S, Neu5NAc. Then, “sialylated lactose” must not be annotated, since it is not an oligosaccharide name.

ix. Mentions indicating part of the oligosaccharide structure

Sometimes oligosaccharides are described thanks to molecule patterns they contain. This information must not be annotated. It can be mentioned as:

- Neu5Gc-containing or Neu5Ac-containing, which are monosaccharides that can be part of the oligosaccharide

The proportion of Neu5Gc-containing oligosaccharides in the acidic oligosaccharide fraction of caprine (64 %) and ovine (94 %) milk was highest (Albrecht_2014_part-2) (Albrecht et al., 2014)

Recent studies indicate that Neu5Gc is found in all mammals except humans and that this is due to a mutation in CMP-sialic acid hydroxylase which occurred in the hominid lineage subsequent to its divergence from the lineage of the great apes. (Urashima_2001_part-1) (Urashima et al., 2001)

– Part of the monosaccharides composing the molecule (Gal(a1-3)Gal). Generally, the terms “residue”, “sequence”, “linked” or “core unit” are used. It must not be annotated since it described only part of the molecule.

Two trisaccharides have a Gal(a1-3)Gal or GalNAc(a1-3)Gal unit at their non reducing ends. (Urashima_2001_part-1) (Urashima et al., 2001)

Due to limitations in enzyme specificities, the (b1 – 3/6)-linked galactotrioses were identified by their respective standards. The (b1 – 4)-linked galactotriose (GU 2·74) was not detected in the milk of any of the animals. (Albrecht_2014_part-2) (Albrecht et al., 2014)

Two of the neutral oligosaccharides have GlcNAc residues at their reducing ends. (Urashima_2001_part-1) (Urashima et al., 2001)

The structures of the silalyl oligosaccharides were: monosialyl lacto-N-neohexaose, monosialyl monofucosyl lacto-N-neohexaose, monosialyl difucosyl lacto-N-neohexaose and disialyl lacto-N-neohexaose. These oligosaccharides contained lacto-N-neohexaose as core units, and one or two a(2-6) linked Neu5Ac, andyor non-reducing a(1-2) linked Fuc. (Urashima_2003b_part-1) (Urashima et al., 2003b)

This sentence is an example of the distinction between part of the structure and molecule names. The mentions “core units” or “linked” indicate that lacto-N-neohexaose does not refer to the molecule here, and thus must not be annotated.

The sugar sequence Gala1-3Galb1-4GlcNAc has been found in mammalian glycoproteins and glycolipids, but free oligosaccharides having this sequence have not previously been reported from natural sources including milk or colostrum. (Urashima_1997_part-1) (Urashima et al., 1997)

x. M/Z

When the authors designate the oligosaccharides with the mass-to-charge m/z (mass divided by charge number of the molecule) ratio, it must not be annotated. In fact, this chemical characteristic is not usable since we cannot relate the m/z to the oligosaccharide.

For example, with BMO, peak 1 was determined to be lacto-N-neotetraose (LNnT) with m/z 710.275 [3Hex + 1HexNAc + H]+ (Tao_2010_part-2) (Tao et al., 2010)

The author gives several synonyms of the molecules, we can annotate the oligosaccharide name as well as its abbreviation and composition, but not the *m/z 710.275* mention.

In our study, the neutral trisaccharide (Hex3, MOs-6, 7) with m/z 529.1739 [M + Na]+ was the most abundant MOs, which comprised more than 50% of total PMOs followed by three common MOs with m/z 636.2346 (MOs-9), 386.1657 (MOs-21), and 677.2612 (MOs-27) with intensities of 5–20% (Table 1). (Cheng_2016_part-2) (Cheng et al., 2016)

Only the composition mention must be annotated in this example.

xi. Molecular mass

The molecular mass is not specified in the oligosaccharide thesaurus. It is necessary to have information about the molecule to assign this mass to a specific oligosaccharide; therefore, it should not be annotated.

Both bovine and porcine milks have SL (molecular mass 635.227) as the most abundant MO (Tao_2010_part-2) (Tao et al., 2010) Here SL includes the 2 isomers (that were not identified separately), it must be annotated. However, molecular mass is not included in the MilkOligoThesaurus, thus we will not be able to find those 2 molecules using this mention.

xii. Succession of synonyms

Sometimes, authors use different formulations to speak about an oligosaccharide (composition, name, detailed structure, m/z...). Each relevant formulation (see the different rules above) must be annotated separately.

The exception is lacto-N-novopentaose I (Gal(b13)[Gal(b1-4)GlcNAc(b1-6)]Gal(b1-4)Glc), which is also found in bovine [7] and equine [8] colostrum and is a prominent constituent of tammar wallaby milk sugars [9] (see below). (Urashima_2001_part-1) (Urashima et al., 2001) Both the name and the structure of the molecule are indicated, they are annotated separately.

xiii. Non-free oligosaccharides

When the milk oligosaccharide described is linked to another molecule (indicated by the –R sign or mentions such as sulphated, phosphate...), it must not be annotated.

Figure 2 shows the structures of saccharides containing the human group A (GalNAc(a1-3)[Fuc(a1-2)]Galb1-R), group B (Gal(a1-3) [Fuc(a1-2)]Galb1-R) and group H (Fuc(a1-2)Galb1-R) antigens, while Figure 3 shows saccharides containing the a-Gal epitope (Gal(a1-3)Gal(b1-4)GlcNAcb1-R). (Urashima_2001_part-1) (Urashima et al., 2001)

With the composition of H3N1S1Su1, peak #15 was predicted to be either the sulphated LSTb or LSTc which are present in rat milk (peak#9 and #10, Table 1). (Li_2021_part-1) (Li et al., 2021) In this example, even the composition is not annotated, since it contains the sulphate ion (Su1), thus it is not part of the milk oligosaccharides.

xiv. Standards

In order to analyze milk and establish the oligosaccharide pattern, some technologies require “standard” molecules. Those are typical milk oligosaccharides, supplied by industrial companies, thus they do not describe mammals’ milk composition and must not be annotated. Generally, they are

found in the methodology section and there is an indication that the molecules have been bought (purchase verb or supplier mention).

Oligosaccharide reference materials (see Table 2for abbreviations), LNFP II, LNFP III, LNFP V, LNT, LNnT, LNnH, LSTc, 2'-FL, and 3-FL, were purchased from Seikagaku Co. (Tokyo, Japan), whereas bovine disialyllactose (Neu5Ac (α2-8)Neu5Ac(α2-3)Gal(β1-4)Glc), 3'-NAc-SL and 6'NAc-SL were obtained from Sigma Co. (St. Louis, MO, USA). B-Tetrasaccharide was isolated from Japanese black bear milk [26], while 3'-NGc-SL was isolated from ovine colostrum [27] (Taufik_2012_part-2) (Taufik et al., 2012)

In this example, the phrase "reference materials" refers to standards, and indicates that those molecules must not be extracted. Most of the standards have been purchased from suppliers while the last ones were directly collected from animals. These oligosaccharides, including those isolated from bear and ovine samples, must not be annotated. It can be hard for an algorithm to understand that the ones extracted from animals are used as standards. Thus, oligosaccharides mentioned in the methodology section must not be extracted.

This pattern was virtually identical to the reference standard of LNFP II. (Taufik_2012_part-2) (Taufik et al., 2012)

This sentence makes no mention of suppliers, and is not found in the methodology but in the results section. However, the word "standard" clearly indicates that LNFP II was used to interpret the result of the study and not to describe the milk oligosaccharide composition.

Since we did not have a quantitative standard, the tetrasaccharide (Hex2-HexNAc-dHex) was quantified using the 2'FL standard curve (Figure 2) assuming equimolar responses for the labeled OS. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The composition must be annotated, as it indicates the presence of isomers. However, the notion tetrasaccharide is too vast and must not be annotated, and neither must 2'FL, which refers to a standard curve used to identify oligosaccharides.

xv. Molecule identifiers specific to an article

Sometimes, authors refer to the oligosaccharides using custom identifiers, as in the example below. These IDs must not be annotated.

Galactotrioses (Gal(α1–3) Gal(b1 – 4)Glc, N6, Gal(b1 – 3)Gal(b1 – 4)Glc, N7, and Gal(b1 –6)Gal(b1 – 4)Glc, N9, with N7 and N9 being identified previously in human milk) are the most abundant structures in the total neutral oligosaccharide pool. (Albrecht_2014_part-2) (Albrecht et al., 2014)

N6, N7... have been assigned to the oligosaccharides so that the author can refer to them in a graph without using their full name (too long for a single graph with multiple molecules). They are excluded from the annotation.

xvi. Unknown molecule names

Whenever the structures detected have not been identified clearly (and can't even be related to isomers), they must not be annotated.

The total number of PMO structures in mature milk increased with parity from gilt to sow (41 vs. 47), with the sialylated PMO known structures of S1441-3, S1043, SF1384 and unknown structures of S1570-1, S1570-2, S1732-3 not being present in gilt milk. (Wei_2018_part-2) (Wei et al., 2018)

It is clearly mentioned that some of the structures encountered have not been assigned to oligosaccharides yet. In addition, the authors use identifiers for the oligosaccharides. Their notations are based on a letter indicating the type of the molecule (neutral, fucosylated...) followed by the m/z found in the study. As these IDs are specific to this article, they must not be annotated.

b. Oligosaccharide types

i. Definition

As mentioned above, oligosaccharides can be gathered in groups based on their monosaccharide composition. We have decided to include 5 major types, listed below and available in the OligosaccharideType-concepts at https://doi.org/10.57745/LFXGFO:

- Neutral
- Fucosylated / Fucosylation
- Sialylated (also known as acidic) / Sialylation
- Type I
- Type II

Since authors use these terms in their text, no thesaurus will be used. However, since sialylated and acidic refer to the same class of molecules, a normalization will be made from acidic to sialylated so that, in the final database, only the mention sialylated will be found.

ii. Other terms used to describe the type of molecules

We have decided to annotate only the 6 adjectives stated above for this entity. However, occasionally authors use other types of words. They must not be annotated.

Ovine and caprine sialyl oligosaccharides, like the bovine ones, contain Neu5Gc as well as Neu5Ac. (Urashima_2001_part-1) (Urashima et al., 2001)

Sialyl oligosaccharides must not be annotated, the term sialyl is too general and can also designate a molecule (Sialyl Lewis).

iii. Term oligosaccharide (OS)

As indicated in the oligosaccharide names section, the term « Oligosaccharides » does not bring precise information. However, when it is surrounded by an oligosaccharide type, it must be annotated with it.

Interestingly, the fucosylated OS (tetrasaccharide A and 2'FL) show much higher variability between the different dog breeds analysed here than the sialyllactoses (3'SL and 6'SL). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The fucosylated term designates a group of oligosaccharides to which all molecules bearing a fucose belong. The other oligosaccharide mentions are molecule names (tetrasaccharide A, 2'FL, 3'SL, 6'SL) or groups of isomers (sialyllactoses).

The predominant fucosylated and sialylated HMOs were 2'-FL and 6'-SL at 40–45 days postpartum and changed to 3-FL and 3'-SL at 200–240 days postpartum. (Liu_2021_part-1) (Liu et al., 2021)

When the pattern "species"-MO(s) follows a type of milk oligosaccharides, it must be included in the type annotation. Thus, it will belong to two entities: "oligosaccharide name" and "oligosaccharide type".

Human milk is commonly considered to be unique when compared with the milk of other species with regard to its high content of complex fucosylated and sialylated lactose-derived oligosaccharides (Kunz_1999_part-1) (Kunz et al., 1999)

This example illustrates discontinuous annotations with "fucosylated oligosaccharides" and "sialylated oligosaccharides" split in the sentence.

In human milk and colostrum, type I saccharides/oligosaccharides dominate: they are present in significantly higher concentrations than type II structures [8–10]. (Taufik_2012_part-2) (Taufik et al., 2012)

As already indicated (4.a.v), general terms used to designate sugars must not be annotated. Thus, in this example, only the term "oligosaccharides" is in a discontinuous annotation with type I and type II.

iv. Types linked to molecules other than oligosaccharides

Just like oligosaccharide names, types can be related to molecules other than oligosaccharides, and must not be annotated. The example below includes sialylated oligosaccharides but also sialylated glycoconjugates, which do not belong to the milk oligosaccharides class. Only the first category must be annotated.

For these reasons, there has recently been an increased focus on the composition of porcine milk, although most analyses have concentrated primarily on protein, lipid, and carbohydrate (Simpson and Nicholas, 2002; Haselhorst et al., 2009; Tao et al., 2010) and paid little attention to the major sialylated glycoconjugate (SiaGC) molecules, namely sialylated glycoprotein (SiaGP), sialylated milk oligosaccharide (Sia-MOS), and gangliosides. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

Parkkinen & Finne (1985) also demonstrated the presence of two phosphorylated sialyl oligosaccharides in bovine colostrum. (Gopal_2000_part-1) (Gopal & Gill, 2000)

The sialyl oligosaccharides mention must not be annotated since the two oligosaccharides belonging to this group are phosphorylated.

v. Monosaccharide mentions

The fucosylated and sialylated types contain monosaccharide Fucose or Sialyl acid respectively. These terms must be excluded from the annotation. In fact, several articles use them to characterize molecules that are not milk oligosaccharides.

The objective of this study was to quantitatively determine the total level of Sia N-acetylneuraminic acid (Neu5Ac), Nglycolylneuraminic acid (Neu5Gc), and ketodeoxynonulosonic acid (KDN) in porcine milk and to compare these levels in gilt and sow milk during lactation. (Jahan_2016_part-1) (Jahan, Wynn & Wang, 2016)

Sia is the sialic acid monosaccharides alone, it must not be annotated

The following new findings are reported: (1) Gilt and sow milk contained significant levels of total Sia, with the highest concentration in colostrum (1,238.5 mg/L), followed by transition milk (778.3 mg/L) and mature milk (347.2 mg/L); (2) during lactation, the majority of Sia was conjugated to Sia-GP (41–46%), followed by Sia-MOS (31–42%) and a smaller proportion in gangliosides (12–28%) (Jahan_2016_part-1) (Jahan, Wynn & Wang, 2016)

Here, "total Sia" includes not only sialylated oligosaccharides but also other molecules linked to sialyl acid, so it is excluded. Conversely, Sia-MOS are sialylated oligosaccharides and must be annotated. However, the quantity mention (5.b.iii) must not be annotated since it is relative to all the sialylated molecules.

Sialic acids (Sia) are key monosaccharide constituents of sialylated glycoproteins (Sia-GP), human sialylated milk oligosaccharide (Sia-MOS), and gangliosides. Human milk sialylated glycoconjugates (Sia-GC) are bioactive compounds known to act as prebiotics and promote neurodevelopment, immune function, and gut maturation in newborns. (Jahan_2016_part-1) (Jahan, Wynn & Wang, 2016)

Sialylated glycoconjugates do not belong to milk oligosaccharides, they are excluded.

The concentrations of total, free, and bound SA were determined using an enzymatic reaction with fluorescence detection (Figure 3). (Mudd_2016_part-2) (Mudd et al., 2016)

SA refers to sialylated with free SA being the monosaccharides sialyl acid while bound SA can refers to sialylated oligosaccharides as well as other compounds that are not oligosaccharides. It must be excluded.

vi. Detailed type of structure

Sometimes, precision is added to the type name with the number of typical structures found in the oligosaccharides. For instance, we can find "mono-sialylated oligosaccharides" which indicates that this molecule belongs to the sialylated group and it has only one sialyl acid. This additional information must not be annotated since we haven't indicated them in the MilkOligoThesaurus (Rumeau et al., 2024).

Peaks #6–#10 were all identified as mono-sialylated oligosaccharides (Fig. 1b and Table 1) while #11–#15 each contain two acidic groups either di-sialylated (#11 and #14) or mono-sialylated and mono-sulphated (#12, #13 and #15). (Li_2021_part-1) (Li et al., 2021)

vii. Subgroups of oligosaccharide types

When there is a more precise division within the oligosaccharide types, it must not be annotated.

For example, the total number of new PMO structures decreased from 25 in colostrum of both sow and gilt milk to 24 and 21 in transitional milk and then further decreased to 19 and 16 in mature milk respectively.... In gilt milk, however, this number decreased from 17 in colostrum to 15 in transitional milk (Fig. 2D). In mature milk, 14 and 11 new sialylated structures were detected in gilt and sow respectively. (Wei_2018_part-2) (Wei et al., 2018)

In this sentence, the oligosaccharides mentioned are new structures. We must not annotate this information.

viii. Combination of different types

As stated in the description of the milk oligosaccharide structure, fucosylated milk oligosaccharides can belong either to neutral or sialylated molecules:

- **Acidic milk oligosaccharides:** contain one or more sialic acid in their structure
  - **Fucosylated acidic milk oligosaccharides**: contain one or more fucose
  - **Non fucosylated acidic milk oligosaccharides**: do not contain any fucose
- **Neutral milk oligosaccharides**: do not contain any sialic acid in their structure
  - **Fucosylated neutral milk oligosaccharides**: contain one or more fucose
  - **Non fucosylated neutral milk oligosaccharides**: do not contain any fucose

Most of the time, authors only mention one of the three classes (fucosylated, sialylated or neutral). However, they sometimes specify whether the fucosylated oligosaccharide is acidic or neutral. In this case, all the indications must be annotated as one entity.

Of the 55 PMO structures identified in this study, 34 were sialylated and 21 were neutral (Tables 1, 2). The neutral sugars included the two fucosylated structures, F566-1 and F566-2, but excluded the sialylated-fucosyl structure SF1384 (Table 1). (Wei_2018_part-2) (Wei et al., 2018)

5. Entities referring to oligosaccharide quantification

OS in milk can be quantified in different ways. Thus, we have decided to split them into 3 entities: absolute quantification, relative quantification (relative abundance), and richness.

a. Absolute quantification

i. Definition

The absolute quantification refers to the total concentration of an oligosaccharide (name or type) in the milk of the species. The unit used is g/L.

Of the dog breeds analyzed here, Schnauzer milk had highest 2'FL levels showing a maximum of approximately 1.2 g/L (ca. 2.5 mM) at around 2–4 days of lactation, thereafter levels dropped to approximately 0.5 g/L (Figure 2C). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The absolute quantification of 2'FL at 2-4 days of lactation must be extracted, however the absolute quantification 0.5 g.L will only be stated without reference to the exact period since we have decided not to take ambiguous words such as "thereafter" (3.c.iii).

ii. Standard deviation

When standard deviation is added to the absolute quantity, it must be annotated in the same entity span.

A decrease ($P < 0.05$) in HMO concentration was observed during the course of lactation for the US mothers, corresponding to 19.3 +/- 2.9 g/L for milk collected on postnatal day 10, decreasing to 8.53 +/- 1.18 g/L on day 120 (repeated measures; n = 14). (Xu_2017_part-1) (Xu et al., 2017)

The absolute quantity and standard deviation are included in the same annotation. Here, HMO is annotated as an oligosaccharide names entity because we have a quantity indication.

iii. Approximation

Sometimes authors use the approximately sign with a number. It must not be annotated.

These results match relative abundance findings in the OS profiling, where sialylated OS decreased throughout lactation and fucosylated OS showed a small, but significant, increase. In general, the most abundant OSs were LNnH (~50 mg/L) and 3'-SL (~30 mg/L), followed by 3-Hex (~20 mg/L), when averaged over lactation. (Mudd_2016_part-2) (Mudd et al., 2016)

iv. Discontinuous annotation with measurement unit

Sometimes, multiple absolute quantities are listed and the author uses only one g/L unit for all numbers. Thus, some of them are separated from the unit, they must be associated to it with through a discontinuous annotation.

The corresponding values for HMOs, EMOs, and bovine milk oligosaccharides (BMOs) present in the corresponding colostrums are about 24, 2.8, and 1 g/L, respectively. (Difilippo_2016_part-2) (Difilippo et al., 2016)

v. Numbers that do not quantify oligosaccharides in samples

Sometimes numbers can be associated with certain oligosaccharides without being a quantification indication. For instance, these numbers can indicate methodology requirements. They must not be annotated. They are generally found in the methodology section of the article.

The limit of quantitation (LOQ) was determined by adding a given amount of HMO, and the limit of detection (LOD) level was calculated with three times the signal-to-noise (S/N) ratio. The LODs were as follows: 0.4 mg/L 2'FL, 5.2 mg/L 3-FL, 2.5 mg/L LNT, 2.8 mg/L LNnT, 1.8 mg/L 3'-SL, and 1.0 mg/L 6'-SL. The LOQs were 9.8 mg/L 2'FL, 19.8 mg/L 3-FL, 19.7 mg/L LNT, 19.9 mg/L LNnT, 9.9 mg/L 3'-SL, and 9.9 mg/L 6'-SL. (Liu_2021_part-2) (Liu et al., 2021)

These numbers indicate the minimum concentration required to detect the different oligosaccharides; they must be excluded.

vi. Interval

Sometimes, the absolute quantity belongs to an interval between two numbers. The interval must be annotated as one entity.

The median value ranges of individual oligosaccharide components in this study were 1013–2891 mg/L 2'-FL, 193–1421 mg/L 3-FL, 314–1478 mg/L LNT, 44–255 mg/L LNnT, 111–241 mg/L 3'-SL, and 23–602 mg/L 6'-SL. (Liu_2021_part-1) (Liu et al., 2021)

In this example, the two numbers are linked by a dash.

Lacto-N-neohexaose (LNnH) and sialyllacto-N-hexaose (S-LNH, isomer not further specified) were absent in some of the colostrum samples (1M, 4M, and 6M) and in mature milk samples (1M# and 3 M#). Among the colostrum samples, the total PMO concentrations ranged from 7.38 to 29.35 g/L. (Difilippo_2016_part-2) (Difilippo et al., 2016)

While in this example, they are linked by "to".

vii. Molarity

When the molar mass is given, we must not annotate this information. In fact, each molecule has a different molar mass, thus in order to make the most of these data, it is required to know molecule detailed information.

On the other hand, 6'SL levels reached about 0.3 to 0.6 g/L (0.5-1 mM) within the first days of lactation and reached a peak level at 5 days of lactation or remained at a rather constant level throughout lactation (Figure 2B). (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The absolute quantification of the oligosaccharides in g/L must be annotated, contrary to the concentration in molar mass.

b. Relative quantification

i. Definition

This entity is a proportion and is expressed as a percentage. It gives information on the proportion of an oligosaccharide or a group of oligosaccharides in the milk of a species. When the relative quantification is compared to the total oligosaccharides, the quantity must be annotated.

Over 30 oligosaccharides (OS) were identified in the milk, with 3'-sialyllactose, lacto-N-tetraose, α1–3,β1–4-d-galactotriose, 2'-fucosyllactose, and 6'-sialyllactose being the most abundant species (accounting for ~70% of the total OS). (Salcedo_2016_part-1) (Salcedo et al., 2016)

This sentence gives us two information. The richness (5.c) of the total oligosaccharides detected and the relative quantity of one oligosaccharide (6'SL) over the total oligosaccharides. We do not annotate the approximately sign.

Both bovine and porcine milks have SL (molecular mass 635.227) as the most abundant MO, which comprised >50% in total PMO (Tao_2010_part-2) (Tao et al., 2010)

We also include the relative quantification of an isomer group. Here, authors have detected a molecule belonging to sialylactose isomers (3'SL and 6'SL), which accounted for more than 50% of the total oligosaccharides.

ii. Discontinuous annotation with percentage unit

Sometimes, multiple relative quantities are listed and the author uses only one percent sign for all numbers. Thus, some of them are separated from the percentage, they must be associated to it with through a discontinuous annotation.

In colostrum, sialylated PMOs are comparable to those in other MMOs, being 60–90, 57, 51, and 20–52% of all MMOs reported for cow, goat, sheep, and horse colostrum, respectively. (Difilippo_2016_part-2) (Difilippo et al., 2016)

This example gathers multiple particularities: relative quantification intervals (this part is explained further away in the paragraph v) as well as discontinuous annotations (60-90, 57 and 51 are separated from the percent sign).

iii. Quantities that are relative to other molecules than the total oligosaccharides

When percentages do not refer to oligosaccharides (4), they must not be annotated.

Based on phenol-sulfuric acid analysis of whole milk, the total saccharide content of mid-lactation milk are: greater galago 6.4%, aye-aye 5.8%, Coquerel's sifaka 4.1% and mongoose lemur 7.9%. Thus, strepsirrhine milks are tentatively estimated to contain about 2.5%, 1.5%, 1.0% and 2.0% oligosaccharide in greater galago, aye-aye, sifaka and mongoose lemur, respectively. (Taufik_2012_part-2) (Taufik et al., 2012)

In the first part of the sentence, percentages give indications about saccharides and are excluded. While in the second part of the extract, percentages refer to the total quantity of oligosaccharides in the milk of those species, thus they must be annotated.

The following new findings are reported: (1) Gilt and sow milk contained significant levels of total Sia, with the highest concentration in colostrum (1,238.5 mg/L), followed by transition milk (778.3 mg/L) and mature milk (347.2 mg/L); (2) during lactation, the majority of Sia was conjugated to Sia-GP (41–46%), followed by Sia-MOS (31–42%) and a smaller proportion in gangliosides (12–28%) (Jahan_2016_part-1) (Jahan, Wynn & Wang, 2016)

The percentage indicated after Sia-MOS (the only molecules that must be annotated here because the others are not milk oligosaccharides) is not the quantification relative to the total oligosaccharides but relative to the total Sia (which do not only include oligosaccharides), it must be excluded.

By 20 days (mature milk), only about 18.3% and 36.1% of the initial content of sialylated MOs was present in sow and gilt milk. In neutral MOs, however, 32.9% and 82.4% of the initial content remained in sow and gilt milk respectively; thus the sialylated MOs were the most variable components of the PMOs. (Wei_2018_part-2) (Wei et al., 2018)

The purpose of the percentage in this extract is to give the evolution pattern of some oligosaccharides. The relative quantification does not refer to the total OS but to a specific fraction measure at a previous time point of the study, it must not be annotated.

iv. Percentages used for other purposes

Sometimes, percentages occur close to molecules names but do not quantify them. Generally, they are mentioned as percentage "scores".

The prominent peaks at 10.4 and 10.8 min corresponded to the triose B(Gal(β1−3)Gal(β1−4)Glc) and triose C(Gal(β1− 6)Gal(β1−4)Glc) with the scores 95% and 97%, respectively (neutral mass: 506.18). (Lee_2016_part-1) (Lee et al., 2016)

The percentages give information on the certainty of the correspondence made between results from the analysis and molecule determination.

We confirmed the presence of a fucosylated oligosaccharide (2'FL) with a match score of 83%. (Lee_2016_part-1) (Lee et al., 2016)

Likewise, 83% quantifies the confidence in the presence of 2'FL.

v. Interval

Sometimes the relative quantification is given as an interval between two values. Thus, the two numbers must be annotated together.

Sixty unique OSs were identified in porcine milk. Neutral OSs were the most abundant at each lactation stage (69–81%), followed by acidic-sialylated OSs (16–29%) and neutral-fucosylated OSs (2–4%). (Mudd_2016_part-1) (Mudd et al., 2016)

vi. Inequality symbols

Sometimes, the number is not precise and authors use inequality symbols (< or >) to indicate this imprecision. These symbols must be annotated with the percentage.

Both bovine and porcine milks have SL (molecular mass 635.227) as the most abundant MO, which comprised >50% in total PMO (Tao_2010_part-2) (Tao et al., 2010)

c. Oligosaccharide richness

i. Definition

This entity gives oligosaccharide quantitative information. It is a number and does not require normalization. It can quantify all the oligosaccharides, a particular oligosaccharide or a group of oligosaccharides.

ii. Total number

As stated in the description of the oligosaccharide name entity (4), the term “oligosaccharide” must be annotated when it gives an indication of the total number of oligosaccharides found in a particular species. Besides the total oligosaccharides, the richness can be used to give the total quantity of a group of oligosaccharides, as well as of one oligosaccharide in particular.

Recently, 39 porcine milk oligosaccharides (PMOs) have been identified, of which 19 are neutral and 20 acidic. (Difilippo_2016_part-2) (Difilippo et al., 2016)

This extract informs us that porcine milk contains at least 39 oligosaccharides. We can also annotate 19 and 20, which characterize the number of oligosaccharide types found in porcine milk.

In contrast to human milk, merely eight neutral fucosylated oligosaccharides were identified in the total neutral oligosaccharide pool from animal milk (Albrecht_2014_part-2) (Albrecht et al., 2014)

The authors give the number of neutral fucosylated oligosaccharides, which is an oligosaccharide type.

iii. Ratio

When the number describes a ratio of one molecule over another, it must not be annotated.

The estimated ratios of oligosaccharide:lactose were: greater galago 1:1.5, aye-aye 1:3, sifaka 1:3 and mongoose lemur 1:3. (Taufik_2012_part-2) (Taufik et al., 2012)

Not only does this sentence relate to lactose, which we have decided to exclude, but it also gives a ratio, which we must not annotate.

iv. Inequality symbols

Sometimes, the number is not precise and authors use inequality symbols (< or >) to indicate this imprecision. These symbols must be annotated with the number.

Approximately 29 different OS were identified in porcine milk, compared to 40 in bovine milk (9) and >130 in human milk (3). (Tao_2010_part-2) (Tao et al., 2010)

v. Comparative and superlative

Adjectives used to quantify oligosaccharides must not be annotated.

Extensive studies from this laboratory (Tao et al. 2008, 2009) and those from other laboratories (Newburg and Neubauer 1995; Gopal and Gill 2000; Mariño et al. 2011) further support that indeed

the most abundant acidic oligosaccharide in bovine milk is sialyllactose. (Aldredge_2013_part-1) (Aldredge et al., 2013)

This extract indicates that the main acidic oligosaccharide is sialyllactose (which includes two isomers). We must annotate the oligosaccharide type as well as the oligosaccharide name to extract the information of their presence in bovine milk, however the mention "most abundant" must not be annotated.

The predominant fucosylated and sialylated HMOs were 2'-FL and 6'-SL at 40–45 days postpartum and changed to 3-FL and 3'-SL at 200–240 days postpartum. (Liu_2021_part-1) (Liu et al., 2021)

Likewise, oligosaccharide types and oligosaccharide names must be annotated to include them in the human milk and associate them to a certain period, however we exclude the fact that they are found in higher quantities at certain period of time.

SL, with m/z at 635.2 (OS-5, Table 1), was clearly the most dominant species (Figure 4, inset), with relative abundances 10 times higher than the next largest OS peak. (Tao_2010_part-2) (Tao et al., 2010)

The mentions "most dominant" and "10 times higher" must be excluded.

Galactotrioses (Gal(a1–3) Gal(b1 – 4)Glc, N6, Gal(b1 – 3)Gal(b1 – 4)Glc, N7, and Gal(b1 –6)Gal(b1 – 4)Glc, N9, with N7 and N9 being identified previously in human milk(7)) are the most abundant structures in the total neutral oligosaccharide pool. (Albrecht_2014_part-2) (Albrecht et al., 2014)

"Most abundant" does not give relevant information, it must not be annotated.

By contrast, the aye-aye and lemur, representing taxa (Chiromyiformes and Lemuriformes) that diverged about 59 mya, shared 8 identified milk oligosaccharides (Table 3). (Taufik_2012_part-2) (Taufik et al., 2012)

This sentence indicates the number of oligosaccharides two species have in common. This information must not be annotated, since we do not compare species between them.

vi. Evolution indication

Terms indicating a time evolution (decrease, increase…) must not be annotated.

Lacto-N-neotetraose decreases significantly as well as lacto-N-neohexaose (Tao_2010_part-2) (Tao et al., 2010)

Oligosaccharides must be annotated to indicate their presence within the species, but no quantities will be associated with them, and neither will the indication of their decrease over lactation.

vii. Statistical numbers

Statistics are sometimes used to give the pattern; they must not be annotated.

In fact, the average number of PMO structures identified in colostrum was 44, while the maximal number identified in certain porcine milk samples reached up to 49 (Fig. 4A). This degree of variation was represented as the Inter-Quartile Range (IQR). The IQR of PMO structural numbers were 4, 7 and 7 in colostrum, transitional and mature milk, respectively (Wei_2018_part-2) (Wei et al., 2018)

6. Relations

This section specifies the relations that must be annotated. Relations are used to link entities so that the extracted information forms a coherent whole (Figure 8). They can be binary (between two entities) or n-ary (between several entities). Each binary relation is described separately, then examples of harder n-ary relations are given. Actually, an entity is frequently part of several relations and binary relations need to be gathered to extract the entire information (the relation between the milk and the species that have produced it must be linked with the oligosaccharides found in the milk). This schema is also available at https://doi.org/10.57745/LFXGFO.

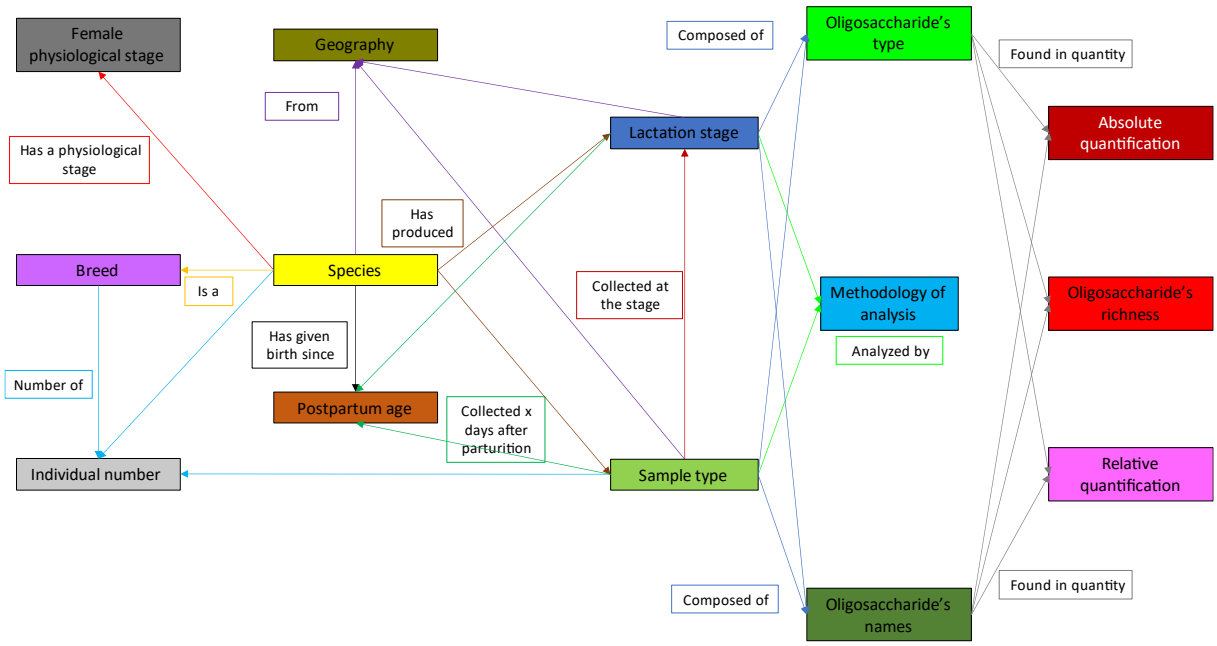

Figure 8: Schematic representation of the relations between entities. Colors used to represent entities correspond to those from the legend; colors of the arrow and relations names do not have significance but only simplify the graph (1 color per relation)

a. Relation "Is a"

This relation gives more detailed information on the studied animals by indicating the breed to which the species belongs. Species is the source of the relation while breed is the target of the relation (Figure 9). Generally, the two entities are close in the sentence.

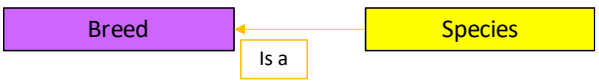

Figure 9: Schematic representation of the relation "is a"

i. One breed per species

When the species came from one breed, the relation to annotate is simple.

Eight milk samples from DutchLandrace sows were collected. (Difilippo_2016_part-2) (Difilippo et al., 2016)

The sows studied are from the breed DutchLandrace

ii. Cross breed species

When a species is a cross breed, we have decided to annotate all the breeds as one entity, this entity must be linked to the species.

Saito et al. (1984) described the chemical structure of two neutral disaccharides in colostrum collected 6 h postparturition from Holstein±Friesian cows. (Gopal_2000_part-1) (Gopal & Gill, 2000)

The cows are cross-breeds of Holstein and Friesian. Both breeds are annotated together as a single entity, which is linked to cows.

iii. Several species mention

When the sentence including the breed mention gathers multiple species synonyms, they must all be linked to the breed's name.

Porcine milk oligosaccharides (PMOs) were analyzed in six colostrum and two mature milk samples from Dutch Landrace sows. (Difilippo_2016_part-2) (Difilippo et al., 2016)

Both the relations between porcine and Dutch Landrace and between sows and Dutch Landrace must be annotated.

Bovine colostrum from Holstein-Friesian cows and porcine colostrum from Landrace pigs were obtained on-site at Teagasc Food Research Centre, Moorepark (Fermoy, Cork, Ireland). (Albrecht_2014_part-2) (Albrecht et al., 2014)

This sentence gathers cross breed species as well as coreferences. Thus, bovine as well as cows must be linked to the Holstein-Friesian cross breed. Moreover, Landrace must be linked not only to porcine but also to pigs.

Sows (n = 10) and gilts (n= 7) of a Landrace, Belgian Landrace, Large White and Duroc cross-breed (Sus scrofa) from the Pig Improvement Company (PIC) (Wei_2018_part-2) (Wei et al., 2018)

The species studied belongs to four breeds (cross-breed). We must annotate them as one entity and link it to the three species mentioned (sow, gilt, and sus scrofa).

b. Relation "From"

This relation indicates where the milk studied came from geographically. We can find the "From" binary relation in three situations (Figure 10):

- Between entities species and Geography
- Between entities Sample type and Geography (Normally the sample type has already been linked to the species from which it was collected)

– Between entities Lactation stage and Geography (Likewise, normally the lactation stage has already been linked to the lactating species)

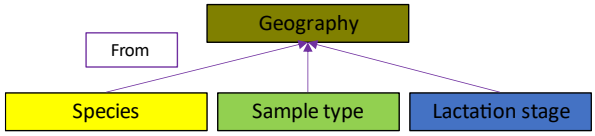

Figure 10: Schematic representation of the relation "From"

i. One geographical mention

When the geographical indication is based on one location, the structure is quite simple.

Seven Yorkshire sows from the University of Illinois Imported Swine Research Laboratory (ISRL) were bred to Yorkshire boars and housed in standard gestation and farrowing crates throughout the study. (Mudd_2016_part-2) (Mudd et al., 2016)

The relation between sows and Illinois is annotated.

ii. Enumerations

When several locations are enumerated, with an inclusion relation between them, they must all be included in the relation individually.

All pigs were obtained from the commercial farrowing shed at the Pig Improvement Company facility (Grong Grong, NSW, Australia) (Wei_2018_part-2) (Wei et al., 2018)

The three location names must be linked to pigs in three distinct relations.

iii. The absence of species mentions in the sentence

Sometimes the species are not mentioned near the geographical entity. In this case, this entity must be linked to a Sample type or Lactation stage entity.

It was validated and applied to milk samples from Malawi (88 individuals; 88 samples from postnatal month 6) (Xu_2017_part-1) (Xu et al., 2017)

No species is mentioned in this sentence; thus, Malawi must be linked to milk.

iv. Choosing between entities

When the entities species and lactation stage or sample type are stated in the sentence, the geographical mention must be linked to the species entity, the latter being in a relation with the lactation stage or sample type entity (see the “have produced” relation).

Mid-lactation milk samples (5ml) were collected from a free-living polar bear with one yearling (16 months old) in Svalbard in the Norwegian Arctic.(Urashima_2003a_part-2) (Urashima et al., 2003b)

Relations between Svalbard and polar bear as well as between Norwegian Arctic and polar bear must be annotated.

v. Coreference of species

When equivalent entities are next to each other, all must be included in separated relations.

Sows (n = 10) and gilts (n= 7) of a Landrace, Belgian Landrace, Large White and Duroc cross-breed (Sus scrofa) from the Pig Improvement Company (PIC) facility at Grong Grong, N.S.W, Australia were the source for collection of all milk samples. (Wei_2018_part-2) (Wei et al., 2018)

The species, already linked to the breeds they belong to, are now linked to the location they originate from. The three locations are included in relations with Sows, Gilts and Sus Scrofa, all referring to pig species.

vi. Indications of location for other purposes

When the location does not refer to the origin of the individual studied, the relation must be excluded.

Milk samples were frozen in air tight vials at −20°C or −80°C and subsequently shipped to the Nutrition Laboratory, Smithsonian National Zoological Park, Washington DC, USA. (Taufik_2012_part-2) (Taufik et al., 2012)

Samples have been moved to the laboratory in the USA but they do not come from this place, the relation with sample type must not be annotated.

c. Relation “Number of”

This relation indicates the number of animals that have been used to characterize the milk (Figure 11).

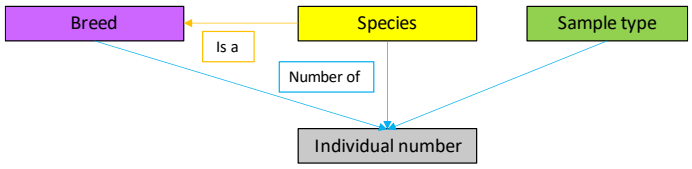

Figure 11: Schematic representation of the relation "Number of"

i. Relation with species

Most of the time, this relationship occurs between species and the Individual number.

Although this suggests that a3galactosyltransferase activity is absent from polar bear mammary glands early in lactation and appears at a later stage, it has to be noted that these results are based on only two animals. (Urashima_2003a_part-2) (Urashima et al., 2003a)

The authors mentioned that their results are based on two polar bears, the number “two” has to be linked to “polar bear”.

ii. Relation with breed

When the species is not mentioned in the sentence, but the breed is, the relation must be annotated between the Individual number and Breed. However, when both species and breed are indicated, only the species entity must be linked to the individual number, while breed is linked to the species (6.a).

Seven Yorkshire sows from the University of Illinois Imported Swine Research Laboratory (ISRL) were bred to Yorkshire boars and housed in standard gestation and farrowing crates throughout the study. (Mudd_2016_part-2) (Mudd et al., 2016)

"Seven" must be linked to "sows" which is itself linked to "Yorkshire". Additionally, the species will be linked to "Illinois", as described in the relation from (6.b).

d. Relation "Has a physiological stage"

This relation indicates to which physiological stage the species belongs, it details whether the species has already given birth or if it is the first time (Figure 12).

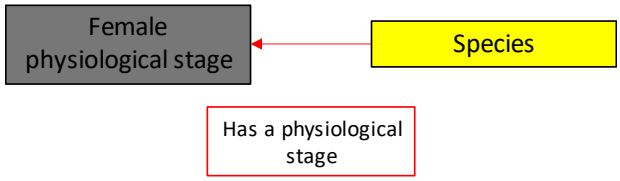

Figure 12: Schematic representation of the relation "Has a physiological stage"

i. Coreference of species terms

When two synonyms of a species are indicated in the same sentence as the physiological stage, they must all be included in separate relations.

Characterization of porcine milk oligosaccharides over lactation between primiparous and multiparous female pigs (Wei_2018_part-1) (Wei et al., 2018)

Primiparous and multiparous are both linked to pigs but also to porcine, they indicate that the milk has been studied on females that have given birth for the first time as well as on females that have already given birth.

ii. Pig specificities

As mentioned, the pig species may be referred to using specific terms indicating whether the female is multiparous or primiparous. These terms are annotated as species entities. However, we may also find female physiological stage entities in addition to these species' terms. In these cases, we must annotate a relation between the species and the physiological stage.

Therefore, the structural diversity of PMOs in the milk of female pigs which had been bred at least once (sow) with those which had been bred for the first time (gilt) were assessed in colostrum (day 1), transitional milk (day 3) and mature milk (day 15–21). (Wei_2018_part-2) (Wei et al., 2018)

"Bred at least once" is linked to "sow", "bred for the first time" is linked to "gilt", and both stages are linked to pigs.

The pigs (Sus scrofa, Belgian Landrace, Large White, Landrace, and Duroc) including gilts (young female pigs that have not farrowed yet; n = 8) and sows (mature female pigs that have bred at least once; n = 22) were included in this study. (Jahan_2016_part-2) (Jahan, Wynn & Wang, 2016)

Likewise, "have not farrowed yet" must be linked to "gilts", as well as "pigs"; "have bred at least once" must be linked to "sows" and "pigs". All these species mention must be linked to the various breeds mentioned.
Additionally, for each type of female (gilt or sow), the number of individuals studied is mentioned: "Gilts" must be linked to "8" while "sows" is linked to "22".

e. Relation "Has given birth since"

This relation indicates the duration since the mother has given birth (Figure 13).

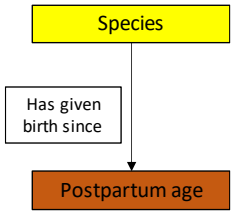

Figure 13: Schematic representation of the relation "Has given birth since"

This relation links Species and Postpartum age entities that are close to each other in the text.

A total of 488 milk samples were collected from 335 healthy lactating mothers: 96 at 0–5 days, 96 at 10–15 days, 104 at 40–45 days, 100 at 200–240 days, and 92 at 300–400 days postpartum. (Liu_2021_part-1) (Liu et al., 2021)

Each of the postpartum age entities must be linked to the species.

f. Relation "Has produced"

The relation connects the species to the analyzed milk it produces (Figure 14).

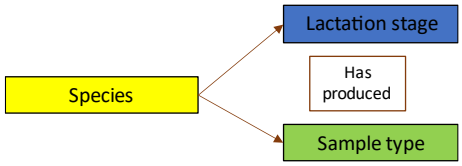

Figure 14: Schematic representation of the relation "Has produced"

i. Relation with sample type

It links Species and Sample type entities to indicate which species has produced the milk studied.

In milk of a mongrel dog the formation of the OS fucosyllactose, 3'sialyllactose and 6'sialyllactose was reported. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

"Dog" must be linked to "milk".

ii. Multiple species for one sample

Sometimes, multiple species can be related to one sample type mention.

The only non-human primates whose milk sugars have been studied appear to be the rhesus monkey and the brown capuchin (Cebus apella). (Urashima_2001_part-1) (Urashima et al., 2001)

Two species are mentioned with only one sample type. Both must be linked to milk, since milk have been studied for both of them.

iii. Species whose milk has not been studied

When multiple species are listed but not all of them have been sampled, only the species from which milk has been collected must be linked to the sample type entity.

All of the assayed strepsirrhine milks included substantial amounts of oligosaccharides. The four species represent all three strepsirrhine infraorders (Lorisiformes [galago], Chiromyiformes [aye-aye] and Lemuriformes [sifaka, lemur]) and four of the seven strepsirrhine families; only Lorisidae [lorises], Cheirogaleidae [dwarf and mouse lemurs] and Lepilemuridae [sportive lemurs] were not sampled. (Taufik_2012_part-2) (Taufik et al., 2012)

Milk has not been collected from lorises, dwarf, mouse lemurs, and sportive lemurs. Thus, only galago, aye-aye, sifaka and lemur have a relation with milk.

iv. Relation with lactation stage

Lactation stage can be used to describe the sample type (6.g) or as a sample. In this last case, a relation must be made between this entity and Species.

The exception is lacto-N-novopentaose I (Gal(b13)[Gal(b1-4)GlcNAc(b1-6)]Gal(b1-4)Glc), which is also found in bovine [7] and equine [8] colostrum and is a prominent constituent of tammar wallaby milk sugars [9]. (Urashima_2001_part-1) (Urashima et al., 2001)

This sentence shows the two relations: bovine and equine species must be linked to colostrum, which is considered as a sample here; tammar wallaby must be linked to milk.

g. Relation "Collected at the stage"

This relation characterizes the sample collected (Figure 15). It indicates the lactation stage during which the sample was collected.

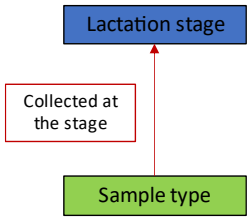

Figure 15: Schematic representation of the relation "Collected at the stage"

i. Details about the sample type

This relation gives details about Sample type by linking it to Lactation stage. It indicates which type of milk it is (the first milk, milk in the middle of lactation period, last milk before weaning...).

The composition of porcine milk oligosaccharides (PMO) was analyzed during early lactation and their relation to piglet gut microbiome was investigated. (Salcedo_pig_part-1) (Salcedo et al., 2016)

The relation made between "milk" and "early lactation" indicates that the milk analyzed from pigs was produced at the beginning of lactation.

To illustrate the relative abundances of the major OS species in porcine milk, three milking points were chosen to represent early (farrow), mid (day 4), and late lactation (day 24) (Tao_2010_part-2) (Tao et al., 2010)

Milk has been collected at different points, corresponding to 3 lactation stages (early and farrow are considered as synonyms in this sentence). They must all be linked to milk.

ii. Lactation stage as a sample

When the lactation stage is used to designate a sample in the same way as the sample type entity, it must not be linked to the sample type entity. In fact, this is a "have produced" relation occurring between lactation stage and species.

Bovine milk normally contains 1–2 g.L of free saccharides other than lactose but larger amounts occur in colostrum (Urashima_2001_part-1) (Urashima et al., 2001)

In this example, authors compare milk and colostrum. Consequently, no relation must be annotated between milk and colostrum, they are both considered as samples collected from cows.

HPLC chromatograms of PMO from colostrum and milk collected at farrowing and days 1, 4, 7, and 24 of lactation from the same sow are shown in Figure 3. (Tao_2010_part-2) (Tao et al., 2010)

This sentence shows the two forms encountered. "Colostrum" is a lactation stage entity used as a sample on its own and must be linked to "sow" (just as "milk" is linked to "sow") but not to "milk". Conversely, "farrowing" is a characterization of the milk, and a relationship is established between these two entities, but not between "farrowing" and "colostrum".

h. Relation "Collected x days after parturition"

This relation gives more details about the samples studied (Figure 16). It indicates the time between the birth and the collection of the samples.

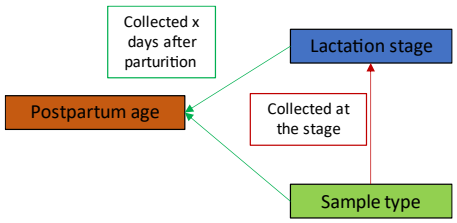

Figure 16: Schematic representation of the relation "Collected x days after parturition"

i. Relation with sample type

The relation is made between Sample type and Postpartum age. In a similar way to the previous relation, it characterizes the sample. Usually, this relation is established when the species entity is not nearby, preventing us from annotating the "has given birth since" relation between Postpartum age and Species.

In this study, we found six oligosaccharides that were present in breast milk from 0 to 400 days postpartum. (Liu_2021_part-1) (Liu et al., 2021)

Postpartum age is linked to milk, which is itself related to the species studied in this article.

ii. Relation with lactation stage

Postpartum age can also be linked to Lactation stage, when the latter is used as a fully-fledged sample and not as a description of a sample type.

The mature milk collected at 4 months post-partum contained lactose, isoglobotriose, 2'-fucosyllactose, B-tetrasaccharide ... However, the mature milk collected at 27 months post-partum contained lactose, isoglobotriose, 3-fucosylisoglobotriose ... (Urashima_2003a_part-2) (Urashima et al., 2003a)

Lactation stages vary among species and have a broad sense (in several species lactation is quite long and thus a lactation stage can last for a long period). The period during which we refer to the milk as mature can last for a long time (4 and 27 months postpartum are included) or can be short. Here postpartum age quantifies this period for the polar bear and thus must be linked to the lactation stage.

iii. Confusion between lactation stage and postpartum age

As already mentioned, Lactation stage and Postpartum age are closely related entities. Sometimes, authors merge both entities to characterize the milk. Therefore, it is necessary to annotate correctly each entity and to link them to the sample type.

Milk was collected from second-parity sows (n =3) at farrowing and on days 1, 4, 7, and 24 of lactation. (Tao_2010_part-1) (Tao et al., 2010)

In this example, lactation stage as well as postpartum age are related to milk (with “collected at the stage” and “collected x days after parturition” relations respectively) but no relations is made between them.

i. Relation “Analyzed by”

This relation indicates how the samples collected have been analyzed (Figure 17).

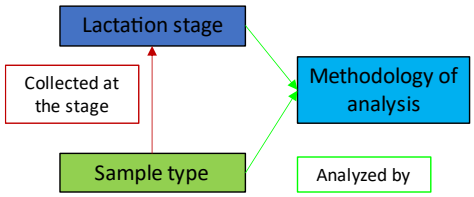

Figure 17: Schematic representation of the relation "Analyzed by"

i. Relation with sample type

Sample type and Methodology are linked by the “analyzed by” relation. Generally, the two entities are found in the same sentence.

A high-performance anion-exchange chromatography with pulsed amperometric detector (HPAEC-PAD) was used to measure HMOs in breast milk as previously described with some modifications. (Liu_2021_part-2) (Liu et al., 2021)

ii. Relation with lactation stage

The relation can also be made between Lactation stage and Methodology when lactation stage is used to designate the analyzed sample.

HPLC chromatograms of PMO from colostrum and milk collected at farrowing and days 1, 4, 7, and 24 of lactation from the same sow are shown in Figure 3. (Tao_2010_part-2) (Tao et al., 2010)

Two binary relations will be made in this example: between “HPLC chromatograms” and “milk” as well as between “HPLC chromatograms” and “colostrum” which is considered as a sample type here.

j. Relation “Composed of”

This relation indicates the oligosaccharides that have been found in the samples studied (Figure 18). It can be made between Oligosaccharide type or Oligosaccharide name and either Sample type or Lactation stage (which frequently refers directly to the sample).

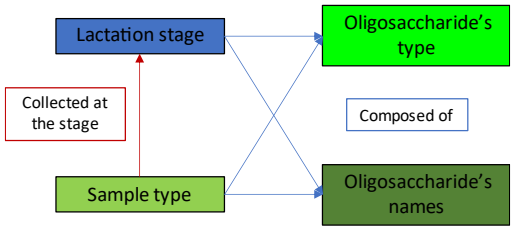

Figure 18: Schematic representation of the relation "Composed of"

These last two entities have been linked to the species studied, which enables us to determine the oligosaccharides that are produced by a specific species.

k. Basic structure

The OS 3'SL, 6'SL and 2'FL were quantified in milk samples of 3 Alaskan huskies (AH, AH-GP, AH-EP), 1 Labrador retriever (LR) and 1 Schnauzer (SCH) using external standard curves of authentic standard compounds. (Rostami_2014_part-2) (Macias Rostami et al., 2014)

The 3 oligosaccharide names are linked to milk. In this sentence, no species is mentioned, however, previously in the article, milk has been linked to dogs. Thus, the final information extracted is that the milk of dogs is composed of 3'SL, 6'SL and 2'FL.

Structural characterization of neutral and acidic oligosaccharides in the milks of strepsirrhine primates: greater galago, aye-aye, Coquerel's sifaka and mongoose lemur (Taufik_2012_part-1) (Taufik et al., 2012)

Relations between both neutral and acidic oligosaccharides must be made with milk. Additionally, all the species named must be linked to milk. In this example, only one sample type is mentioned, while 2 oligosaccharide types and 4 species are named. Thus, "milks" will have multiple relations. In this case it is not an issue since all species have the same oligosaccharide composition. However, a few examples further away illustrate the complexity of these types of structures.

i. Absence of sample type entity

Sometimes the entities are not close to each other in the sentence. If they are only slightly distant, the relation must still be annotated.

Six OSs were present in all the samples analyzed across lactation (LDFH-I, 2'-FL, LNFP-I, LNnH, 3 Hex, 3'-SL; Figure 2), while LDFT was present only in colostrum samples (data not shown). The OS LNT, 2Hex-1HexNAc, 6'-SLN, and 6'-SL were not quantifiable in any sample due to concentrations being below detectable levels.
Total quantified OS concentrations in porcine milk decreased throughout lactation. (Mudd_2016_part-2) (Mudd et al., 2016)

"LDFT" is in close proximity to the lactation stage, and the relation is made easily. "milk" is further away, but even so, it should be linked to the oligosaccharides. In addition, the colostrum entity will be linked to "porcine", otherwise we will not know the species in which this oligosaccharide was found (6.k.ii).

ii. Lack of specific species

When the species are not clearly identified, we must not extract their information.

Three possible structures including Neu5Acα2-3Galβ1-3 Gal ß1-4Glc, Neu5Acα2-3(Galβ1-6)Galβ1-4Glc and Galβ1-3(Neu5Acα2-6)Galβ1-4Glc have been reported in non-human mammalian milk. (Urashima_2001_part-1) (Urashima et al., 2001)

We know that there are 3 oligosaccharides in the milk of non-human mammals. However, this group of individuals (non-human mammals) is too vast and we have decided not to annotate it, thus if we link the oligosaccharides to the milk, we will not know to which mammal this milk belongs.

The total oligosaccharide pools from the milk of all the domestic animals were composed of approximately 80 –90 % acidic oligosaccharides (Table 4) ... twenty-nine neutral, forty-five sialylated and three phosphorylated structures were identified in animal milk (Tables 1 – 3). (Albrecht_2014_part-2) (Albrecht et al., 2014)

Likewise, no relations will be annotated, since it gives us an indication of the composition of a larger group: domestic animals, that we have decided not to annotate.

l. Relation “Found in quantity”

This relation quantifies the oligosaccharides found in the milk of the species studied (Figure 19). Several entities can be linked in this relation:

- Oligosaccharide’s type and Absolute quantification
- Oligosaccharide’s type and Relative quantification
- Oligosaccharide’s type and Oligosaccharide’s richness
- Oligosaccharide’s name and Absolute quantification
- Oligosaccharide’s name and Relative quantification
- Oligosaccharide’s name and Oligosaccharide’s richness

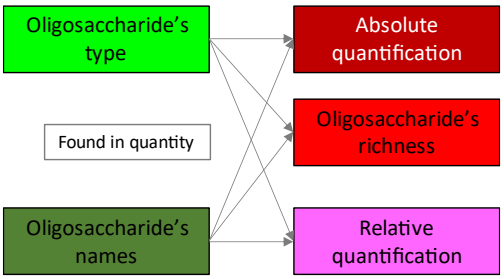

Figure 19: Schematic representation of the relation "Found in quantity"

i. Simple structure

There were >55 sialylated structures identified in human milk and 35 sialylated structures were reported in bovine milk (Wei_2018_part-2) (Wei et al., 2018)

“Human” and “milk” are linked by the “have produced” relation; “sialylated” is linked to “milk” thanks to the “composed of” relation and “>55” is linked to “sialylated” with the relation “found in quantity”.

This process is the same for “bovine”. This sentence is quite simple since each species is clearly separated.

Total fucosylation is around 1% of the OS abundance in porcine colostrum and even less in mature porcine milk. (Tao_2010_part-2) (Tao et al., 2010)

2 lactation stages are mentioned. Both contain the fucosylated oligosaccharides, however quantification only concerns one of these stages. “1%” is linked to “fucosylation”, which in turn is connected to “colostrum”. This last entity is linked to “porcine”. “Fucosylation” must also be linked to “mature milk”, however no quantification is indicated. Consequently, “1%” must be related to the right lactation stage.

ii. Complex structure

Sometimes, a lot of information is given and it can be complex to correctly link the different entities.

In colostrum, there was essentially no difference in the total number of neutral (19 each), sialylated (33 vs. 32) and fucosylated (2 vs 2) MOs between the sow and gilt; The major changes in different numbers of MO structures between sow and gilt milk were found in transitional milk (53 vs. 48). Interestingly, sialylated MO structures were responsible for most of these differences (28 vs. 32), while there was essentially no change in the number of neutral PMOs (from 17 to 18) in transitional milk;. (Wei_2018_part-2) (Wei et al., 2018)

Oligosaccharide’s richness entities are linked to Oligosaccharide type entities, which are themselves linked to Lactation stage entities, which in turn are connected to Species entities. We have complete information on the quantity of specific types of oligosaccharides found in a specific milk of the species. It is necessary to make n-ary relations so that the correct number of oligosaccharides is associated to the corresponding species (53 oligosaccharides are present in the transitional milk of the sow, while there are only 48 in the transitional milk of the gilt).

iii. Quantification of unwanted oligosaccharides

When oligosaccharides we have decided not to extract are quantified, the entities will be annotated but not the relation.

For example, the total number of new PMO structures decreased from 25 in colostrum of both sow and gilt milk to 24 and 21 in transitional milk and then further decreased to 19 and 16 in mature milk respectively.... In gilt milk, however, this number decreased from 17 in colostrum to 15 in transitional milk (Fig. 2D). In mature milk, 14 and 11 new sialylated structures were detected in gilt and sow respectively. (Wei_2018_part-2) (Wei et al., 2018)

In this sentence, the oligosaccharides mentioned are new structures. Thus, the quantity does not represent the total richness of the oligosaccharide types (sialylated, neutral...) but a group among this total richness. Consequently, we must not extract this information.

iv. Multiple quantifications

On postnatal day 180, the total concentration of HMOs in Malawi milk samples from secretors (6.46 +/- 1.74 mg/mL) was higher (P < 0.05) than that in samples from nonsecretors (5.25 +/- 2.55 mg/mL) (Xu_2017_part-1) (Xu et al., 2017)

Two measures are given for the absolute quantification of HMOs in human milk. They are specific to a genetic type of individual, which we have decided not to extract (2.a.xi). Consequently, both measures must be linked to "HMOs", which are linked to "milk" (related to Malawi and postnatal day 180).

v. Absence of the structure

When an oligosaccharide is not present in a species milk, the oligosaccharide name will be annotated, but not the relation. Sometimes, the sentence is structured in a complex form, with multiple species mentioned and only a few having the oligosaccharide.

Compared with up to 70% fucosylation of hMO, there is little or no fucosylated OS in the milk of domestic animals, including bovine (5). (Tao_2010_part-2) (Tao et al., 2010)

In the following sentence, the fucosylated OS must not be linked to bovine milk since they are not present. We do annotate the oligosaccharides as well as bovine milk but we do not link them together. The absence concerns more species than only bovines, however they are referred to as "domestic animals", which do not correspond to a species entity.

vi. Inclusion of several molecules in the quantification

When there is more than one oligosaccharide included in the quantification it must not be annotated. We make an exception for oligosaccharides included in a larger group, which we have decided to annotate (the 5 oligosaccharides types, isomers groups).

Together, 3- and 6-sialyl-lactose account for more than 50 % of the total oligosaccharides present in bovine colostrum. (Gopal_2000_part-1) (Gopal & Gill, 2000)

50% include two oligosaccharides, we must not annotate this information.

Six OSs comprised 60% of the total (2 Hex-1 Neu5Ac or sialyllactose; 3 Hex-1 HexNAc; 3 Hex; 4 Hex-1 HexNAc; 4 Hex-2 HexNAc; and 4 Hex-2 HexNAc-1 Neu5Ac), with 4 Hex-2 HexNAc (25–33%) being the most abundant at each stage of lactation. (Mudd_2016_part-2) (Mudd et al., 2016)

60% includes six oligosaccharides, thus we must not annotate this relative quantity. On the contrary, 25-33% quantifies only one group of isomers (same composition), then this information must be annotated.

vii. Absence

Relations denoting the absence of a certain type of oligosaccharides must not be annotated. Thus, it is important to be careful not to include these oligosaccharides in the milk of the species studied.

Only one NeuGc-linked OS, a(2-6)NeuGclactosamine, was present in porcine colostrum. It was undetectable in mature porcine milk. (Tao_2010_part-2) (Tao et al., 2010)

The oligosaccharide must only be linked to porcine colostrum but not to mature porcine milk.

We detected type I saccharides (LNFP II and LNT) in aye-aye milk but not in greater galago, Coquerel's sifaka or mongoose lemur milk. (Taufik_2012_part-2) (Taufik et al., 2012)

We must link type I, LNFP II and LNT to the milk of aye-aye but not link them to the other species in which they have not been detected.

Isoglobotriose was consistently a dominant saccharide in the mature milk of three species of bear (polar bear, Japanese black bear and Ezo brown bear), but was not found in this sample of polar bear colostrum. (Urashima_2003a_part-2) (Urashima et al., 2003a)

Isoglobotriose is found in the mature milk of the three species, but not in the colostrum of one of these species. All entities are annotated but the relation between isoglobotriose and colostrum must not be annotated.

References

Albrecht S, Lane JA, Mariño K, Al Busadah KA, Carrington SD, Hickey RM, Rudd PM. 2014. A comparative study of free oligosaccharides in the milk of domestic animals. *British Journal of Nutrition* 111:1313–1328. DOI: 10.1017/S0007114513003772.

Aldredge DL, Geronimo MR, Hua S, Nwosu CC, Lebrilla CB, Barile D. 2013. Annotation and structural elucidation of bovine milk oligosaccharides and determination of novel fucosylated structures. *Glycobiology* 23:664–676. DOI: 10.1093/glycob/cwt007.

Balogh R, Jankovics P, Béni S. 2015. Qualitative and quantitative analysis of N-acetyllactosamine and lacto-N-biose, the two major building blocks of human milk oligosaccharides in human milk samples by high-performance liquid chromatography-tandem mass spectrometry using a porous graphitic carbon column. *Journal of chromatography. A* 1422:140–146. DOI: 10.1016/j.chroma.2015.10.006.

Bode L. 2012. Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology* 22:1147–1162. DOI: 10.1093/glycob/cws074.

Cheng L, Xu Q, Yang K, He J, Chen D, Du Y, Yin H. 2016. Annotation of porcine milk oligosaccharides throughout lactation by hydrophilic interaction chromatography coupled with quadruple time of flight tandem mass spectrometry. *Electrophoresis* 37:1525–1531. DOI: 10.1002/elps.201500471.

Difilippo E, Pan F, Logtenberg M, Willems R (H. AM, Braber S, Fink-Gremmels J, Schols HA, Gruppen H. 2016. Milk Oligosaccharide Variation in Sow Milk and Milk Oligosaccharide Fermentation in Piglet Intestine. *Journal of Agricultural and Food Chemistry* 64:2087–2093. DOI: 10.1021/acs.jafc.6b00497.

Gopal PK, Gill HS. 2000. Oligosaccharides and glycoconjugates in bovine milk and colostrum. *British Journal of Nutrition* 84:69–74. DOI: 10.1017/S0007114500002270.

Jahan M, Wynn PC, Wang B. 2016. Molecular characterization of the level of sialic acids N-acetylneuraminic acid, N-glycolylneuraminic acid, and ketodeoxynonulosonic acid in porcine milk during lactation. *Journal of dairy science* 99:8431–8442. DOI: 10.3168/jds.2016-11187.

Kunz C, Rudloff S, Schad W, Braun D. 1999. Lactose-derived oligosaccharides in the milk of elephants: comparison with human milk. *British Journal of Nutrition* 82:391–399. DOI: 10.1017/S0007114599001798.

Lee H, Cuthbertson DJ, Otter DE, Barile D. 2016. Rapid Screening of Bovine Milk Oligosaccharides in a Whey Permeate Product and Domestic Animal Milks by Accurate Mass Database and Tandem Mass Spectral Library. *Journal of agricultural and food chemistry* 64:6364–6374. DOI: 10.1021/acs.jafc.6b02039.

Leong A, Liu Z, Almshawit H, Zisu B, Pillidge C, Rochfort S, Gill H. 2019. Oligosaccharides in goats' milk-based infant formula and their prebiotic and anti-infection properties. *The British journal of nutrition* 122:441–449. DOI: 10.1017/S000711451900134X.

Li J, Jiang M, Zhou J, Ding J, Guo Z, Li M, Ding F, Chai W, Yan J, Liang X. 2021. Characterization of rat and mouse acidic milk oligosaccharides based on hydrophilic interaction chromatography coupled with electrospray tandem mass spectrometry. *Carbohydrate polymers* 259:117734. DOI: 10.1016/j.carbpol.2021.117734.

Liu S, Cai X, Wang J, Mao Y, Zou Y, Tian F, Peng B, Hu J, Zhao Y, Wang S. 2021. Six Oligosaccharides' Variation in Breast Milk: A Study in South China from 0 to 400 Days Postpartum. *Nutrients* 13:4017. DOI: 10.3390/nu13114017.

Macias Rostami S, Bénet T, Spears J, Reynolds A, Satyaraj E, Sprenger N, Austin S. 2014. Milk oligosaccharides over time of lactation from different dog breeds. *PloS one* 9:e99824. DOI: 10.1371/journal.pone.0099824.

Mudd AT, Salcedo J, Alexander LS, Johnson SK, Getty CM, Chichlowski M, Berg BM, Barile D, Dilger RN. 2016. Porcine Milk Oligosaccharides and Sialic Acid Concentrations Vary Throughout Lactation. *Frontiers in Nutrition* 3. DOI: 10.3389/fnut.2016.00039.

Rudloff S, Kuntz S, Rasmussen S, Roggenbuck M, Sprenger N, Kunz C, Sangild P, Bering S. 2019. Metabolism of Milk Oligosaccharides in Preterm Pigs Sensitive to Necrotizing Enterocolitis. *Frontiers in Nutrition* 6. DOI: 10.3389/fnut.2019.00023.

Rumeau M, Fenaille F, Girard A, Loux V, Ba M, Nédellec C, Deléger L, Bossy R, Aubin S, Knudsen C, Combes S. 2024. MilkOligoThesaurus, a dataset of mammalian milk oligosaccharide synonyms. *Data in Brief* 54:110404. DOI: 10.1016/j.dib.2024.110404.

Salcedo J, Frese SA, Mills DA, Barile D. 2016. Characterization of porcine milk oligosaccharides during early lactation and their relation to the fecal microbiome. *Journal of dairy science* 99:7733–7743. DOI: 10.3168/jds.2016-10966.

Sprenger N, Tytgat HLP, Binia A, Austin S, Singhal A. 2022. Biology of human milk oligosaccharides: From basic science to clinical evidence. *Journal of Human Nutrition and Dietetics* 35:280–299. DOI: 10.1111/jhn.12990.

Tao N, Ochonicky KL, German JB, Donovan SM, Lebrilla CB. 2010. Structural Determination and Daily Variations of Porcine Milk Oligosaccharides. *Journal of Agricultural and Food Chemistry* 58:4653–4659. DOI: 10.1021/jf100398u.

Taufik E, Fukuda K, Senda A, Saito T, Williams C, Tilden C, Eisert R, Oftedal O, Urashima T. 2012. Structural characterization of neutral and acidic oligosaccharides in the milks of strepsirrhine primates: greater galago, aye-aye, Coquerel's sifaka and mongoose lemur. *Glycoconjugate Journal* 29:119–134. DOI: 10.1007/s10719-012-9370-9.

Thurl S, Munzert M, Boehm G, Matthews C, Stahl B. 2017. Systematic review of the concentrations of oligosaccharides in human milk. *Nutrition Reviews* 75:920–933. DOI: 10.1093/nutrit/nux044.

Trevisi P, Luise D, Won S, Salcedo J, Bertocchi M, Barile D, Bosi P. 2020. Variations in porcine colostrum oligosaccharide composition between breeds and in association with sow maternal performance. *Journal of Animal Science and Biotechnology* 11. DOI: 10.1186/s40104-020-0430-x.

Ujihara T, Kentaro Y. 2022. The current manufacturing process and business development of human milk oligosaccharide products. *Glycoforum* 25:A7.

Urashima T, Kusaka Y, Nakamura T, Saito T, Maeda N, Messer M. 1997. Chemical characterization of milk oligosaccharides of the brown bear, Ursus arctos yesoensis. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1334:247–255. DOI: 10.1016/S0304-4165(96)00101-8.

Urashima T, Nagata H, Nakamura T, Arai I, Saito T, Imazu K, Hayashi T, Derocher AE, Wiig O. 2003a. Differences in oligosaccharide pattern of a sample of polar bear colostrum and mid-lactation milk. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 136:887–896. DOI: 10.1016/j.cbpc.2003.09.001.

Urashima T, Nakamura T, Yamaguchi K, Munakata J, Arai I, Saito T, Lydersen C, Kovacs KM. 2003b. Chemical characterization of the oligosaccharides in milk of high Arctic harbour seal (Phoca vitulina vitulina). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 135:549–563. DOI: 10.1016/S1095-6433(03)00130-2.

Urashima T, Saito T, Nakamura T, Messer M. 2001. Oligosaccharides of milk and colostrum in non-human mammals. *Glycoconjugate Journal* 18:357–371. DOI: 10.1023/A:1014881913541.

Varki A, Cummings RD, Aebi M, Packer NH, Seeberger PH, Esko JD, Stanley P, Hart G, Darvill A, Kinoshita T, Prestegard JJ, Schnaar RL, Freeze HH, Marth JD, Bertozzi CR, Etzler ME, Frank M, Vliegenthart JF, Lütteke T, Perez S, Bolton E, Rudd P, Paulson J, Kanehisa M, Toukach P, Aoki-Kinoshita KF, Dell A, Narimatsu H, York W, Taniguchi N, Kornfeld S. 2015. Symbol Nomenclature for Graphical Representations of Glycans. *Glycobiology* 25:1323–1324. DOI: 10.1093/glycob/cwv091.

Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (eds.). 2022. *Essentials of Glycobiology*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.

Wei J, Wang ZA, Wang B, Jahan M, Wang Z, Wynn PC, Du Y. 2018. Characterization of porcine milk oligosaccharides over lactation between primiparous and multiparous female pigs. *Scientific Reports* 8:4688. DOI: 10.1038/s41598-018-23025-x.

Xu G, Davis JC, Goonatilleke E, Smilowitz JT, German JB, Lebrilla CB. 2017. Absolute Quantitation of Human Milk Oligosaccharides Reveals Phenotypic Variations during Lactation. *The Journal of Nutrition* 147:117–124. DOI: 10.3945/jn.116.238279.


RÉPUBLIQUE
FRANÇAISE
Liberté
Égalité
Fraternité
INRAE